WEST Search History

PG/U502/41117

DATE: Friday, May 02, 2003

Set Name side by side	Query	Hit Count	Set Name result set
DB=USPT,PGPB,JPAB,EPAB,DWPI,TDBD; PLUR=YES; OP=OR			
L1	tanimoto adj coefficient	21	L1
L2	penrose adj distance	0	L2
L3	mahalanobis	414	L3
L4	mahalanobia adj distance	0	L4
L5	jarvis-patrick	4	L5
L6	jarvis adj patrick adj cluster$	0	L6
L7	mahalanobis adj distance	342	L7
L8	L1 and (protein$ or polypeptide$ or peptide$)	20	L8
L9	protein$ or polypeptide$ or peptide$	372984	L9
L10	l1 and l9	20	L10
L11	l3 and l9	35	L11
L12	l7 and l9	24	L12
L13	l5 and l9	3	L13
DB=USPT; PLUR=YES; OP=OR			
L14	l10	8	L14
DB=USPT,PGPB,JPAB,EPAB,DWPI,TDBD; PLUR=YES; OP=OR			
L15	l10 or l11 or l12 or l13	56	L15
DB=USPT; PLUR=YES; OP=OR			
L16	L15	24	L16
DB=PGPB; PLUR=YES; OP=OR			
L17	L15	32	L17
DB=JPAB,EPAB,DWPI,TDBD; PLUR=YES; OP=OR			
L18	L15	0	L18
DB=USPT,PGPB; PLUR=YES; OP=OR			
L19	@pd<20011221	6874686	L19
L20	l16 and l19	17	L20
DB=USPT; PLUR=YES; OP=OR			
L21	l16 and l19	17	L21
DB=PGPB; PLUR=YES; OP=OR			
L22	l17 and l19	2	L22

Displayed twice w/o PR.

END OF SEARCH HISTORY

* * * * * * * * * * Welcome to STN International * * * * * * * * * *

NEWS 1 Web Page URLs for STN Seminar Schedule - N. America NEWS 2 Apr 08 "Ask CAS" for self-help around the clock NEWS 3 Jun 03 New e-mail delivery for search results now available NEWS 4 Aug 08 PHARMAMarketLetter(PHARMAML) - new on STN NEWS 5 Aug 19 Aquatic Toxicity Information Retrieval (AQUIRE) now available on STN NEWS 6 Aug 26 Sequence searching in REGISTRY enhanced NEWS 7 Sep 03 JAPIO has been reloaded and enhanced NEWS 8 Sep 16 Experimental properties added to the REGISTRY file NEWS 9 Sep 16 CA Section Thesaurus available in CAPLUS and CA NEWS 10 Oct 01 CASREACT Enriched with Reactions from 1907 to 1985 NEWS 11 Oct 24 BEILSTEIN adds new search fields NEWS 12 Oct 24 Nutraceuticals International (NUTRACEUT) now available on STN NEWS 13 Nov 18 DKILIT has been renamed APOLLIT NEWS 14 Nov 25 More calculated properties added to REGISTRY NEWS 15 Dec 04 CSA files on STN NEWS 16 Dec 17 PCTFULL now covers WP/PCT Applications from 1978 to date NEWS 17 Dec 17 TOXCENTER enhanced with additional content NEWS 18 Dec 17 Adis Clinical Trials Insight now available on STN NEWS 19 Jan 29 Simultaneous left and right truncation added to COMPENDEX, ENERGY, INSPEC NEWS 20 Feb 13 CANCERLIT is no longer being updated NEWS 21 Feb 24 METADEX enhancements NEWS 22 Feb 24 PCTGEN now available on STN NEWS 23 Feb 24 TEMA now available on STN NEWS 24 Feb 26 NTIS now allows simultaneous left and right truncation NEWS 25 Feb 26 PCTFULL now contains images NEWS 26 Mar 04 SDI PACKAGE for monthly delivery of multifile SDI results NEWS 27 Mar 19 APOLLIT offering free connect time in April 2003 NEWS 28 Mar 20 EVENTLINE will be removed from STN NEWS 29 Mar 24 PATDPAFULL now available on STN NEWS 30 Mar 24 Additional information for trade-named substances without structures available in REGISTRY NEWS 31 Apr 11 Display formats in DGENE enhanced NEWS 32 Apr 14 MEDLINE Reload NEWS 33 Apr 17 Polymer searching in REGISTRY enhanced NEWS 34 Apr 21 Indexing from 1947 to 1956 being added to records in CA/CAPLUS NEWS 35 Apr 21 New current-awareness alert (SDI) frequency in WPIDS/WPINDEX/WPIX NEWS 36 Apr 28 RDISCLOSURE now available on STN

NEWS EXPRESS April 4 CURRENT WINDOWS VERSION IS V6.01a, CURRENT MACINTOSH VERSION IS V6.0b(ENG) AND V6.0Jb(JP), AND CURRENT DISCOVER FILE IS DATED 01 APRIL 2003 NEWS HOURS STN Operating Hours Plus Help Desk Availability NEWS INTER General Internet Information NEWS LOGIN Welcome Banner and News Items NEWS PHONE Direct Dial and Telecommunication Network Access to STN NEWS WWW CAS World Wide Web Site (general information)

Enter NEWS followed by the item number or name to see news on that specific topic.

All use of STN is subject to the provisions of the STN Customer agreement. Please note that this agreement limits use to scientific research. Use for software development or design or implementation of commercial gateways or other similar uses is prohibited and may result in loss of user privileges and other penalties.

* * * * * * * * * * * * * STN Columbus * * * * * * * * * * * * *

FILE 'HOME' ENTERED AT 12:36:16 ON 02 MAY 2003

=> caplus
CAPLUS IS NOT A RECOGNIZED COMMAND
The previous command name entered was not recognized by the system.
For a list of commands available to you in the current file, enter
"HELP COMMANDS" at an arrow prompt (=>).

=> file caplus
COST IN U.S. DOLLARS SINCE FILE TOTAL ENTRY SESSION
FULL ESTIMATED COST 0.21 0.21

FILE 'CAPLUS' ENTERED AT 12:36:28 ON 02 MAY 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

FILE COVERS 1907 - 2 May 2003 VOL 138 ISS 19
FILE LAST UPDATED: 1 May 2003 (20030501/ED)
This file contains CAS Registry Numbers for easy and accurate substance identification.

=> s (polypeptide? or peptide? or protein?)/bi,ab 122730 POLYPEPTIDE?/BI 111260 POLYPEPTIDE?/AB 371018 PEPTIDE?/BI 286650 PEPTIDE?/AB 1777224 PROTEIN?/BI 1454496 PROTEIN?/AB
L1 1996186 (POLYPEPTIDE? OR PEPTIDE? OR PROTEIN?)/BI,AB

=> s (classif? or cluster?)/bi,ab 135281 CLASSIF?/BI 122591 CLASSIF?/AB 209658 CLUSTER?/BI 187380 CLUSTER?/AB
L2 340424 (CLASSIF? OR CLUSTER?)/BI,AB

=> s l1 and l2
L3 43130 L1 AND L2

=> s align?/bi,ab 86557 ALIGN?/BI 78322 ALIGN?/AB
L4 86557 ALIGN?/BI,AB

=> s l3 and l4
L5 1597 L3 AND L4

=> s (sequenc? (5w) analy?)/bi,ab 668419 SEQUENC?/BI 568798 SEQUENC?/AB 2269045 ANALY?/BI 1178311 ANALY?/AB

L6 22544 (SEQUENC? (5W) ANALY?)/BI,AB

=> s l5 and l6
L7 170 L5 AND L6

=> s l7 not 2003/py 331777 2003/PY
L8 167 L7 NOT 2003/PY

=> s l8 not 2002/py 1057064 2002/PY

L9 134 L8 NOT 2002/PY

=> d l9 1-134 bib ab

L9 ANSWER 1 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2002:176910 CAPLUS
DN 137:196603
TI Noncoding RNA gene detection using comparative sequence analysis
AU Rivas, Elena; Eddy, Sean R.
CS Howard Hughes Medical Inst. and Dep. Genetics, Washington Univ. Sch. Medicine, St. Louis, MO, USA
SO BMC Bioinformatics [online computer file] (2001), 2, No pp. given CODEN: BBMIC4; ISSN: 1471-2105 URL: http://www.biomedcentral.com/content/pdf/1471-2105-2-8.pdf
PB BioMed Central Ltd.
DT Journal; (online computer file)
LA English
AB Noncoding RNA genes produce transcripts that exert their function without ever producing proteins . Noncoding RNA gene sequences do not have strong statistical signals, unlike protein coding genes. A reliable general purpose computational gene finder for noncoding RNA genes has been elusive. Results: We describe a comparative sequence anal. algorithm for detecting novel structural RNA genes. The key idea is to test the pattern of substitutions obsd. in a pairwise alignment of two homologous sequences. A conserved coding region tends to show a pattern of synonymous substitutions, whereas a conserved structural RNA tends to show a pattern of compensatory mutations consistent with some base-paired secondary structure. We formalize this intuition using three probabilistic "pair-grammars": a pair stochastic context free grammar modeling alignments constrained by structural RNA evolution, a pair hidden Markov model modeling alignments constrained by coding sequence evolution, and a pair hidden Markov model modeling a null hypothesis of position-independent evolution. Given an input pairwise sequence alignment (e.g. from a BLASTN comparison of two related genomes) we classify the alignment into the coding, RNA, or null class according to the posterior probability of each class. Conclusions: We have implemented this approach as a program, QRNA, which we consider to be a prototype structural noncoding RNA gene finder. Tests suggest that this approach detects noncoding RNA genes with a fair degree of reliability.
RE.CNT 43 THERE ARE 43 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 2 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2002:45230 CAPLUS
DN 136:396559
TI Browsing gene banks for Fe2S2 ferredoxins and structural modeling of 88 plant-type sequences : an analysis of fold and function
AU Bertini, Ivano; Luchinat, Claudio; Provenzani, Alessandro; Rosato, Antonio; Vasos, Paul R.
CS Centro di Risonanze Magnetiche, Department of Chemistry, University of Florence, Sesto Fiorentino, 50019, Italy
SO Proteins: Structure, Function, and Genetics (2001), Volume Date 2002, 46(1), 110-127 CODEN: PSFGEY; ISSN: 0887-3585
PB Wiley-Liss, Inc.
DT Journal
LA English
AB One-hundred-and-seventy-nine sequences of Fe2S2 ferredoxins and ferredoxin precursors were identified in and retrieved from currently available protein and cDNA databases. On the basis of their cluster -binding patterns, these sequences were divided into three groups: those contg. the CX4CX2CXnC pattern (plant-type ferredoxins), those with the CX5CX2CXnC pattern (adrenodoxins), and those with a different pattern. These three groups contain, resp., 139, 36, and 4 sequences. After excluding ferredoxin precursors in the first group, two subgroups were identified, again based on their cluster -binding patterns: 88 sequences had the CX4CX2CX29C pattern, and 29 had the CX4CX2CXmC (m 29) pattern. The structures of the 88 ferredoxins with the CX4CX2CX29C pattern were modeled based on the available exptl. structures of nine proteins within this same group. The modeling procedure was tested by building structural models for the ferredoxins with known structures. The models resulted, on av., in being within 1 .ANG. of the backbone root-mean-square deviation from the corresponding exptl. structures. In addn., these structural models were shown to be of high quality by using assessment procedures based on energetic and stereochem. parameters. Thus, these models formed a reliable structural database for this group of ferredoxins, which is meaningful within the framework of current structural genomics efforts. From the anal. of the structural database generated it was obsd. that the secondary structural elements and the overall three-dimensional structures are maintained throughout the superfamily. In particular, the residues in the hydrophobic core of the protein were either absolutely conserved or conservatively substituted. In addn., certain solvent-accessible charged groups, as well as hydrophobic groups, were conserved to the same degree as the core residues. The patterns of conservation of exposed residues identified the regions of the protein that are crit. for its function in electron transfer. An extensive anal. of protein - protein interactions is now possible. Some conserved interactions between residues have been identified and related to structural and/or functional features. All this information could not be obtained from the analyses of the primary sequences alone. Finally, the anal. of the sequences of the related subgroup featuring the CX4CX2CXmC cluster -binding pattern in the light of the structural and functional insights provided by the inspection of the mentioned structural database affords some hints on the functional features of ferredoxins belonging to this subgroup.
RE.CNT 48 THERE ARE 48 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 3 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2002:30347 CAPLUS
DN 136:196139
TI The (.beta..alpha.)8 glycosidases: sequence and structure analyses suggest distant evolutionary relationships
AU Nagano, Nozomi; Porter, Craig T.; Thornton, Janet M.
CS Biomolecular Structure and Modelling Group, Biochemistry & Molecular Biology Department, University College London, London, WC1E 6BT, UK

SO Protein Engineering (2001), 14(11), 845-855 CODEN: PRENE9; ISSN: 0269-2139
PB Oxford University Press
DT Journal
LA English
AB There are currently at least nine distinct glycosidase sequence families which are all known to adopt a TIM barrel fold. To explore the relationships between these enzymes and their evolution, comprehensive sequence and structure comparisons were performed, generating four distinct clusters . The first cluster , S1, comprises the .alpha.-amylase related enzymes, all with the retention mechanism (axialaxial). The second cluster , S2, included two functional subgroups, one composed of various kinds of glucosidases all with the retention mechanism (equatorial.fwdarw.equatorial) (the so-called 4/7 superfamily), and the other subgroup including the .beta.-amylases with the inversion mechanism (axial.fwdarw.equatorial). The third cluster , S3, with the retention mechanism (equatorial.fwdarw.equatorial), could be subdivided, based on the catalytic residues and mechanisms, into two functional subgroups: the chitinase group, catalyzed by two acidic residues on the C-termini of .beta.-4 and .beta.-6, and the hevamine group, using two acidic residues on the C-termini of .beta.-4 for catalysis. The fourth cluster , S4, is composed of chitobiase with the retention mechanism (equatorial.fwdarw.equatorial). These clusters are compared with the sequence families derived by Henrissat and coworkers. PSI-BLAST profiles and multiple- alignments of tertiary structures suggest that S1 and S2 are distantly related, as are S3 and S4, which have N-acetylated substrates. This work highlights the difficulties of untangling distant evolutionary relationships in ubiquitous folds such as the TIM barrel.
RE.CNT 37 THERE ARE 37 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 4 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2002:95 CAPLUS
DN 136:364262
TI Bioinformatic tools for DNA/ protein sequence analysis , functional assignment of genes and protein classification
AU Rehm, B. H. A.
CS Institut fuer Mikrobiologie, Westfalischen Wilhelms-Universitaet, Muenster, 48149, Germany
SO Applied Microbiology and Biotechnology (2001), 57(5-6), 579-592 CODEN: AMBIDG; ISSN: 0175-7598
PB Springer-Verlag
DT Journal; General Review
LA English
AB A review. The development of efficient DNA sequencing methods has led to the achievement of the DNA sequence of entire genomes from (to date) 55 prokaryotes, 5 eukaryotic organisms and 10 eukaryotic chromosomes. Thus, an enormous amt. of DNA sequence data is available and even more will be forthcoming in the near future. Anal. of this overwhelming amt. of data requires bioinformatic tools in order to identify genes that encode functional proteins or RNA. This is an important task, considering that even in the well-studied Escherichia coli more than 30% of the identified open reading frames are hypothetical genes. Future challenges of genome sequence anal. will include the understanding of gene regulation and metabolic pathway reconstruction including DNA chip technol., which holds tremendous potential for biomedicine and the biotechnol. prodn. of valuable compds. The overwhelming vol. of information often confuses scientists. This review intends to provide a guide to choosing the most efficient way to analyze a new sequence or to collect information on a gene or protein of interest by applying current publicly available databases and Web services. Recently developed tools that allow functional assignment of genes, mainly based on sequence similarity of the deduced amino acid sequence. Using the currently available and increasing biol. databases will be discussed.
RE.CNT 71 THERE ARE 71 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 5 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:921197 CAPLUS
DN 137:106322
TI Genome trees constructed using five different approaches suggest new major bacterial clades
AU Wolf, Yuri I.; Rogozin, Igor B.; Grishin, Nick V.; Tatusov, Roman L.; Koonin, Eugene V.
CS National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, 20894, USA
SO BMC Evolutionary Biology [online computer file] (2001), 1, No pp. given CODEN: BEBMCG; ISSN: 1471-2148 URL: http://www.biomedcentral.com/1471-2148/1/8
PB BioMed Central Ltd.
DT Journal; (online computer file)
LA English
AB Background: The availability of multiple complete genome sequences from diverse taxa prompts the development of new phylogenetic approaches, which attempt to incorporate information derived from comparative anal. of complete gene sets or large subsets thereof. Such attempts are particularly relevant because of the major role of horizontal gene transfer and lineage-specific gene loss, at least in the evolution of prokaryotes. Results: Five largely independent approaches were employed to construct trees for completely sequenced bacterial and archaeal genomes: i) presence-absence of genomes in clusters of orthologous genes; ii) conservation of local gene order (gene pairs) among prokaryotic genomes; iii) parameters of identity distribution for probable orthologs; iv) anal. of concatenated alignments of ribosomal proteins ; v) comparison of trees constructed for multiple protein families. All constructed trees support the sepn. of the two primary prokaryotic domains, bacteria and archaea, as well as some terminal bifurcations within the bacterial and archaeal domains. Beyond these obvious groupings, the trees made with different methods appeared to differ substantially in terms of the relative contributions of phylogenetic relationships and similarities in gene repertoires caused by similar life styles and horizontal gene transfer to the tree topol. The trees based on presence-absence of genomes in orthologous clusters and the trees based on conserved gene pairs appear to be strongly affected by gene loss and horizontal gene transfer. The trees based on identity distributions for orthologs and particularly the tree made of concatenated ribosomal protein sequences seemed to carry a stronger phylogenetic signal. The latter tree supported three potential high-level bacterial clades: i) Chlamydia-Spirochetes, ii) Thermotogales-Aquificales (bacterial hyperthermophiles), and ii) Actinomycetes-Deinococcales- Cyanobacteria. The latter group also appeared to join the low-GC Gram-pos. bacteria at a deeper tree node. These new groupings of bacteria were supported by the anal. of alternative topologies in the concatenated ribosomal protein tree using the Kishino-Hasegawa test and by a census of the topologies of 132 individual groups of orthologous proteins . Addnl., the results

of this anal. put into question the sister-group relationship between the two major archaeal groups, Euryarchaeota and Crenarchaeota, and suggest instead that Euryarchaeota might be a paraphyletic group with respect to Crenarchaeota. Conclusions: We conclude that, the extensive horizontal gene flow and lineage-specific gene loss notwithstanding, extension of phylogenetic anal. to the genome scale has the potential of uncovering deep evolutionary relationships between prokaryotic lineages.
RE.CNT 51 THERE ARE 51 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 6 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:836784 CAPLUS
DN 136:65897
TI Dimerization of G- Protein -Coupled Receptors
AU Dean, Mark K.; Higgs, Christopher; Smith, Richard E.; Bywater, Robert P.; Snell, Christopher R.; Scott, Paul D.; Upton, Graham J. G.; Howe, Trevor J.; Reynolds, Christopher A.
CS Department of Biological Sciences Central Campus, University of Essex, Colchester Essex, CO4 3SQ, UK
SO Journal of Medicinal Chemistry (2001), 44(26), 4595-4614 CODEN: JMCMAR; ISSN: 0022-2623
PB American Chemical Society
DT Journal
LA English
AB The evolutionary trace (ET) method, a data mining approach for detg. significant levels of amino acid conservation, has been applied to over 700 aligned G- protein -coupled receptor (GPCR) sequences. The method predicted the occurrence of functionally important clusters of residues on the external faces of helixes 5 and 6 for each family or subfamily of receptors; similar clusters were obsd. on helixes 2 and 3. The probability that these clusters are not random was detd. using Monte Carlo techniques. The cluster on helixes 5 and 6 is consistent with both 5,6-contact and 5,6-domain swapped dimer formation; the possible equivalence of these two types of dimer is discussed because this relates to activation by homo- and heterodimers. The observation of a functionally important cluster of residues on helixes 2 and 3 is novel, and some possible interpretations are given, including heterodimerization and oligomerization. The application of the evolutionary trace method to 113 aligned G- protein sequences resulted in the identification of two functional sites. One large, well-defined site is clearly identified with adenyl cyclase, .beta./.gamma. and regulator of G- protein signaling (RGS) binding. The other G- protein functional site, which extends from the ras-like domain onto the helical domain, has the correct size and electrostatic properties for GPCR dimer binding. The implications of these results are discussed in terms of the conformational changes required in the G-protein for activation by a receptor dimer. Further, the implications of GPCR dimerization for medicinal chem. are discussed in the context of these ET results.
RE.CNT 188 THERE ARE 188 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 7 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:827917 CAPLUS
DN 137:43804
TI Protein sequence-structure space and resultant data redundancy in the protein data bank
AU Shindyalov, I. N.; Bourne, P. E.
CS San Diego Supercomputer Center, University of California San Diego, La Jolla, CA, 92093, USA
SO METMBS '01, Proceedings of the International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences, Las Vegas, NV, United States, June 25-28, 2001 (2001), 139-145. Editor(s): Valafar, Faramarz. Publisher: CSREA Press, Athens, Ga. CODEN: 69BZQV
DT Conference
LA English
AB A study of sequence-structure space and resultant data redundancy has been performed using the Combinatorial Extension (CE) algorithm for detg. structural alignment and BLAST for detg. sequence similarity. Significant clusters in sequence-structure space assocd. with recurrent structures (convergent evolution) and protein superfamilies (divergent evolution) have been described. These observations have been compared to the SCOP classification of protein domains that define similar features. Both methods indicate an enormous redundancy of data in the Protein Data Bank (PDB), and hence a need in defining representative (non-redundant) sets of proteins esp. for use in various computational analyses. We propose here an approach for building representative sets using combined sequence and structure similarity criterion with addnl. conditions requiring adequate representation of proteins excluded from the set. Representative sets are updated on a weekly basis and available from http://cl.sdsc.edu/nr.html.
RE.CNT 13 THERE ARE 13 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 8 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:825383 CAPLUS
DN 137:43798
TI Markovian domain fingerprinting: Statistical segmentation of protein sequences
AU Bejerano, Gill; Seldin, Yevgeny; Margalit, Hanah; Tishby, Naftali
CS School of Computer Science & Engineering, The Hebrew University, Jerusalem, 91904, Israel
SO Bioinformatics (2001), 17(10), 927-934 CODEN: BOINFP; ISSN: 1367-4803
PB Oxford University Press
DT Journal
LA English
AB Characterization of a protein family by its distinct sequence domains is crucial for functional annotation and correct classification of newly discovered proteins . Conventional Multiple Sequence Alignment (MSA) based methods find difficulties when faced with heterogeneous groups of proteins . However, even many families of proteins that do share a common domain contain instances of several other domains, without any common underlying linear ordering. Ignoring this modularity may lead to poor or even false classification results. An automated method that can analyze a group of proteins into the sequence domains it contains is therefore highly desirable. We apply a novel method to the problem of protein domain detection. The method takes as input an unaligned group of protein sequences. It segments them and clusters the segments into groups sharing the same underlying statistics. A Variable Memory Markov (VMM) model is built using a Prediction Suffix Tree (PST) data structure for each group of segments. Refinement is achieved by letting the PSTs compete over the segments, and a deterministic annealing framework infers the no. of underlying PST models while avoiding many inferior solns. We show that regions of similar statistics correlate well with protein sequence domains,

by matching a unique signature to each domain. This is done in a fully automated manner, and does not require or attempt an MSA. Several representative cases are analyzed. We identify a protein fusion event, refine an HMM superfamily classification into the underlying families the HMM cannot sep., and detect all 12 instances of a short domain in a group of 396 sequences.
RE.CNT 22 THERE ARE 22 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 9 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:819989 CAPLUS
DN 136:336889
TI Molecular cloning and sequence analysis of stearoyl-CoA desaturase in milkfish, Chanos chanos
AU Hsieh, S. L.; Liao, W. L.; Kuo, C. M.
CS Institute of Fisheries Science, National Taiwan University, Taipei, 106, Taiwan
SO Comparative Biochemistry and Physiology, Part B: Biochemistry & Molecular Biology (2001), 130B(4), 467-477
CODEN: CBPBB8; ISSN: 1096-4959
PB Elsevier Science Inc.
DT Journal
LA English
AB Stearoyl-CoA desaturase (E.C. 1.14.99.5) is a key enzyme in the biosynthesis of polyunsatd. fatty acids and the maintenance of the homeoviscous fluidity of biol. membranes. The stearoyl-CoA desaturase cDNA in milkfish (Chanos chanos) was cloned by RT-PCR and RACE, and it was compared with the stearoyl-CoA desaturase in cold-tolerant teleosts, common carp and grass carp. Nucleotide sequence anal. revealed that the cDNA clone has a 972-bp open reading frame encoding 323 amino acid residues. Alignments of the deduced amino acid sequence showed that the milkfish stearoyl-CoA desaturase shares 79% and 75% identity with common carp and grass carp, and 63%-64% with other vertebrates such as sheep, hamsters, rats, mice, and humans. Like common carp and grass carp, the deduced amino acid sequence in milkfish well conserves three histidine cluster motifs (one HXXXXH and two HXXHH) that are essential for catalysis of stearoyl-CoA desaturase activity. However, RT-PCR anal. showed that stearoyl-CoA desaturase expression in milkfish is detected in the tissues of liver, muscle, kidney, brain, and gill, and more expression sites were found in milkfish than in common carp and grass carp. Phylogenic relationships among the deduced stearoyl-CoA desaturase amino acid sequence in milkfish and those in other vertebrates showed that the milkfish stearoyl-CoA desaturase amino acid sequence is phylogenetically closer to those of common carp and grass carp than to other higher vertebrates.
RE.CNT 47 THERE ARE 47 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 10 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:779285 CAPLUS
DN 136:33603
TI Iron-sulfur flavoprotein (Isf) from Methanosarcina thermophila is the prototype of a widely distributed family
AU Zhao, Tong; Cruz, Francisco; Ferry, James G.
CS Department of Biochemistry and Molecular Biology, Eberly College of Science, The Pennsylvania State University, University Park, PA, 16802-4500, USA
SO Journal of Bacteriology (2001), 183(21), 6225-6233
CODEN: JOBAAY; ISSN: 0021-9193
PB American Society for Microbiology
DT Journal
LA English
AB A total of 35 homologs of the iron-sulfur flavoprotein (Isf) from Methanosarcina thermophila were identified in databases. All three domains were represented, and multiple homologs were present in several species. An unusually compact cysteine motif ligating the 4Fe-4S cluster in Isf is conserved in all of the homologs except two, in which either an aspartate or a histidine has replaced the second cysteine in the motif. A phylogenetic anal. of Isf homologs identified four subgroups, two of which were supported by bootstrap data. Three homologs from metabolically and phylogenetically diverse species in the Bacteria and Archaea domains (Af3 from Archaeoglobus fulgidus, Cd1 from Clostridium difficile, and Mj2 from Methanococcus jannaschii) were overproduced in Escherichia coli. Each homolog purified as a homodimer, and the UV-visible absorption spectra were nearly identical to that of Isf. After reconstitution with iron, sulfide, and FMN the homologs contained six to eight nonheme iron atoms and 1.6 to 1.7 FMN mols. per dimer, suggesting that two 4Fe-4S or 3Fe-4S clusters and two FMN cofactors were bound to each dimer, which is consistent with Isf data. Homologs Af3 and Mj2 were reduced by CO in reactions catalyzed by cell ext. of acetate-grown M. thermophila, but Cd1 was not. Homologs Af3 and Mj2 were reduced by CO in reactions catalyzed by A. fulgidus and M. jannaschii cell exts. Cell ext. of Clostridium thermoaceticum catalyzed CO redn. of Cd1. Our database sequence analyses and biochem. characterizations indicate that Isf is the prototype of a family of iron-sulfur flavoproteins that occur in members of all three domains.
RE.CNT 26 THERE ARE 26 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 11 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:617266 CAPLUS
DN 135:177722
TI Protein Functional Sub-type Analysis and Prediction from Sequence Alignments
IN Hannenhalli, Sridhar; Russell, Robert B.
PA Smith Klein Beecham Corporation, USA; Smithkline Beecham Plc
SO Jpn. Kokai Tokkyo Koho, 64 pp. CODEN: JKXXAF
DT Patent
LA Japanese
FAN.CNT 1 PATENT NO. KIND DATE APPLICATION NO. DATE
---------- ---- -------- ---------- --------
PI JP 2001229201 A2 20010824 JP 2000-329217 20001027
PRAI US 1999-162456P P 19991029
AB This invention presents a new approach for analyzing and predicting subtypes from protein sequence alignments . Given a multiple sequence alignment and a classification of different subtypes (e.g. differences in enzyme specificity), the profile difference method exploits the differences between hidden Markov model profiles to highlight positions on the sequences that are most discerning of each subtype. The method is insensitive to conservative substitutions, and tolerates missing data by combining alignments with amino acid exchange matrixes via the construction of an HMM (Eddy, 1998). For new sequences known to be homologous to an existing family, but of unknown subtype, the method can exploit the known subtype classifications and assocd. profiles to predict subtype. The increasing no. and diversity of protein sequence families requires new methods to define and predict details regarding function. Here, we present a method for anal. and prediction of functional sub-types from multiple protein sequence alignments . Given an alignment and set of proteins grouped

into sub-types according to some definition of function, such as enzymic specificity, the method identifies positions that are indicative of functional differences by comparison of sub-type specific sequence profiles, and anal. of positional entropy in the alignment . Alignment positions with significantly high positional relative entropy correlate with those known to be involved in defining sub-types for nucleotidyl cyclases, protein kinases, lactate/malate dehydrogenases and trypsin-like serine proteases. We highlight new positions for these proteins that suggest addnl. expts. to elucidate the basis of specificity. The method is also able to predict sub-type for unclassified sequences. We assess several variations on a prediction method, and compare them to simple sequence comparisons. For assessment, we remove close homologs to the sequence for which a prediction is to be made (by a sequence identity above a threshold). This simulates situations where a protein is known to belong to a protein family, but is not a close relative of another protein of known sub-type. Considering the four families above, and a sequence identity threshold of 30 %, our best method gives an accuracy of 96 % compared to 80 % obtained for sequence similarity and 74 % for BLAST. We describe the derivation of a set of sub-type groupings derived from an automated parsing of alignments from PFAM and the SWISSPROT database, and use this to perform a large-scale assessment. The best method gives an av. accuracy of 94 % compared to 68 % for sequence similarity and 79 % for BLAST. We discuss implications for exptl. design, genome annotation and the prediction of protein function and protein intra-residue distances.

L9 ANSWER 12 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:605673 CAPLUS
DN 135:315532
TI DELPHI: A pattern-based method for detecting sequence similarity
AU Floratos, A.; Rigoutsos, I.; Parida, L.; Gao, Y.
CS First Genetic Trust, Inc., Lyndhurst, NJ, 07071, USA
SO IBM Journal of Research and Development (2001), 45(3/4), 455-473 CODEN: IBMJAE; ISSN: 0018-8646
PB International Business Machines Corp.
DT Journal
LA English
AB We describe DELPHI, a new computational tool for identifying sequence similarity between a query sequence and a database of proteins . Use is made of a set of patterns obtained from the underlying database through a one-time computation. The patterns are subsequently matched against every query sequence presented to the system. A pattern matched by a region of the query pinpoints a potential local similarity between that region and all of the database sequences also matching that pattern. In a final step, all such local similarities are examd. more closely by aligning and scoring the corresponding query and database regions. By prudently choosing a set of patterns, the method can be used to discover weak but biol. important similarities. We provide a no. of examples using both classified and unclassified proteins that corroborate this claim.
RE.CNT 32 THERE ARE 32 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 13 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:600618 CAPLUS
DN 136:258020
TI Partial sequence analysis of the actin gene and its potential for studying the phylogeny of Candida species and their teleomorphs
AU Daniel, Heide-M.; Sorrell, Tania C.; Meyer, Wieland
CS Molecular Mycology Laboratory, ICPMR, The University of Sydney/Westmead Hospital, Westmead, NSW 2145, Australia
SO International Journal of Systematic and Evolutionary Microbiology (2001), 51(4), 1593-1606 CODEN: ISEMF5; ISSN: 1466-5026
PB Society for General Microbiology
DT Journal
LA English
AB The actin gene has been studied as a potential phylogenetic marker for selected members of the anamorphic genus Candida and seven related teleomorphic genera (Debaryomyces, Issatchenkia, Kluyveromyces, Saccharomyces and Pichia from the Saccharomycetaceae; Clavispora and Metschnikowia from the Metschnikowiaceae). The nucleotide sequences of 36 fungal taxa were analyzed with respect to their mol. evolution and phylogenetic relationships. A total of 460 bp (47%) of the coding 979 bp were variable and 396 bp (40%) of these were found to be phylogenetically informative. Further anal. of the sequences showed that the genic G+C contents were higher than the nuclear G+C contents for most of the taxa. A strong pos. correlation was found between G+C content over all codon positions and third positions. First and second codon positions were considered to be independent of the genic G+C content. The expected transition/transversion bias was detected only for third positions. Pairwise comparisons of transitional and transversional changes (substitutions) with total percentage sequence divergences revealed that the third position transitions showed no satn. for ingroup comparisons. A sp. wt.ing scheme was set up, combining codon-position wts. with change-frequency wts. to enable the inclusion of distant outgroup taxa. Parsimony analyses of the investigated taxa showed four groups, three of which corresponded to major clusters that had been established previously in Candida by rDNA anal. Interrelationships among the species groups in this heterogeneous anamorphic genus were detd. The polyphyletic origin of the selected Candida species and their close assocns. with several ascomycete genera were verified and known anamorph/teleomorph pairs confirmed. The actin gene was established as a valuable phylogenetic marker with the particular advantage of an unambiguous alignment .
RE.CNT 64 THERE ARE 64 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 14 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:512179 CAPLUS
DN 136:161887
TI Massive sequence comparisons as a help in annotating genomic sequences
AU Louis, Alexandra; Ollivier, Emmanuelle; Aude, Jean-Christophe; Risler, Jean-Loup
CS Laboratoire Genome et Informatique, Universite de Versailles, Versailles, 78035, Fr.
SO Genome Research (2001), 11(7), 1296-1303 CODEN: GEREFS; ISSN: 1088-9051
PB Cold Spring Harbor Laboratory Press
DT Journal
LA English
AB An all-by-all comparison of all the publicly available protein sequences from plants has been performed, followed by a clusterization process. Within each of the 1064 resulting clusters -contg. sequences that are orthologous as well as paralogous-the sequences have been submitted to a pyramidal classification and their domains delineated by an automated procedure a la PRODOM. This process provides a

means for easily checking for any apparent inconsistency in a cluster , for example, whether one sequence is shorter or longer than the others, one domain is missing, etc. In such cases, the alignment of the DNA sequence of the gene with that of a close homologous protein often reveals (in 10% of the clusters) probable sequencing errors (leading to frameshifts) or probable wrong intron/exon predictions.
RE.CNT 28 THERE ARE 28 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 15 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:477182 CAPLUS
DN 136:161855
TI A model for phylogenetic inference using structural and chemical covariates
AU Tavare, Simon; Adams, Dean C.; Fedrigo, Olivier; Naylor, Gavin J. P.
CS Departments of Biological Sciences, Mathematics and Preventative Medicine, University of Southern California, Los Angeles, CA, 90089, USA
SO Pacific Symposium on Biocomputing 2001, Mauna Lani, HI, United States, Jan. 3-7, 2001 (2001), 215-225. Editor(s): Altman, Russ B. Publisher: World Scientific Publishing Co. Pte. Ltd., Singapore, Singapore. CODEN: 69BLFC
DT Conference
LA English
AB We investigated whether or not evolutionary change in DNA sequence data was homogeneous across different classes of base pairs. DNA sequences for eight protein -coding mitochondrial genes were obtained for 38 vertebrate taxa from GenBank. Each nucleotide site in the alignment was classified according to a no. of covariates, including its codon position, genetic code degeneracy, and hydrophobicity. The evolutionary transition matrix for each base was estd. by tracing implied character changes under parsimony on a known phylogenetic tree. Canonical variates analyses of the inferred transition matrixes were performed for each gene to det. whether or not different classes of bases behaved similarly. We found five distinct clusters of transition matrixes that could be roughly defined by combinations of codon position and degeneracy. This pattern was consistent among all genes. A stochastic model of rate variation based on the interaction of the covariates was developed to assess the statistical significance of the clusters . The five-group classification was found to explain significantly more sequence variation than did a codon only classification , a codon degeneracy classification , or a codon and degeneracy classification . The same five-group classification was found for all genes tested, suggesting a common process underlying the mol. evolution of the mitochondrial genome. These results confirm that there are classes of base pairs that evolve differently, and suggest that models of sequence evolution that incorporate covariate information may be useful in developing nucleotide substitution models that more accurately reflect evolutionary history.
RE.CNT 24 THERE ARE 24 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 16 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:465920 CAPLUS
DN 135:353297
TI Sequences and topology
AU Gerstein, Mark; Honig, Barry
CS Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, 06520, USA
SO Current Opinion in Structural Biology (2001), 11(3), 327-329 CODEN: COSBEF; ISSN: 0959-440X
PB Elsevier Science Ltd.
DT Journal; General Review
LA English
AB A commentary, with refs., on reviews presented by various authors in the accompanying papers (ibid 11:330-376). The Editorial overview provides a summary of the reviews that discuss how to take the vast amt. of biosequence information, such genome sequences, three-dimensional structure of proteins and expression data sets, and translate it into meaningful information regarding the function of a product. The reviews also include a wide variety of computational approaches, such as sequence and structure alignment and anal., gene-expression clustering and biophys. anal. The reviews further touch on genome annotation, integration of expression information, fold assignments, structural alignment and the understanding of protein - protein interactions. A common thread that runs through many of the reviews is the use of clustering to define biol. parts and then the use of these parts as frameworks for data integration and anal.
RE.CNT 6 THERE ARE 6 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 17 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:457184 CAPLUS
DN 135:192957
TI Are Red Algae Plants? A Critical Evaluation of Three Key Molecular Data Sets
AU Stiller, John W.; Riley, Jennifer; Hall, Benjamin D.
CS Department of Biology, East Carolina University, Greenville, NC, 27858, USA
SO Journal of Molecular Evolution (2001), 52(6), 527-539 CODEN: JMEVAU; ISSN: 0022-2844
PB Springer-Verlag New York Inc.
DT Journal
LA English
AB Whether red algae are related to green plants has been debated for over a century. Features present due to their shared photosynthetic habit have been interpreted as support for an evolutionary sisterhood of the two groups but, until very recently, characters endogenous to the host cell have provided no reliable indication of such a relationship. In this investigation, we examine three mol. data sets that have provided key evidence of a possible relationship between green plants and red algae. Analyses of an expanded alignment of DNA-dependent RNA polymerase II largest subunit sequences indicate that their support for independent origins of rhodophytes and chlorophytes is not the result of long-branch attraction, as has been proposed elsewhere. Differences in the pol II C-terminal domain, an essential component of plant mRNA transcription, also suggest different host cell ancestors for the two groups. In contrast, concatenated sequences of two groups of mitochondrial genes, those encoding subunits of NADH-dehydrogenase as well as cytochrome c oxidase subunits plus apocytochrome B, appear to cluster red algal and green plant sequences together because both groups have evolved relatively slowly and share a super-abundance of ancestral positions. Finally, analyses of elongation factor 2 sequences demonstrate a strong phylogenetic signal favoring a rhodophyte/chlorophyte sister relationship, but that signal is restricted to a contiguous segment comprising approx. half of the EF2 gene. These results argue for great caution in the interpretation of phylogenetic analyses of ancient evolutionary events but, in combination, indicate that there is no emerging consensus

from mol. data supporting a sister relationship between red algae and green plants.
RE.CNT 58 THERE ARE 58 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 18 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:350154 CAPLUS
DN 135:238845
TI Picasso: Generating a covering set of protein family profiles
AU Heger, Andreas; Holm, Liisa
CS Structural Genomics Group, EMBL-EBI, Cambridge, CB10 1SD, UK
SO Bioinformatics (2001), 17(3), 272-279 CODEN: BOINFP; ISSN: 1367-4803
PB Oxford University Press
DT Journal
LA English
AB Evolutionary classification leads to an economical description of protein sequence data because attributes of function and structure are inherited in protein families. This paper presents Picasso, a procedure for deriving a minimal set of protein family profiles that cover all known protein sequences. Picasso starts from highly overlapping sequence neighborhoods revealed by all-on-all pairwise Blast alignment . Overlaps are reduced by merging sequences or parts of sequences into multiple alignments . For max. unification, the multiple alignments must reach into the twilight zone of sequence similarity. Sensitive and selective profile-profile comparison allows unification down to about 15% pairwise sequence identity. Families unified through a short conserved sequence motif are assocd. with multiple full-length alignments describing different subfamilies. Domains that are mobile modules are identified based on their assocn. with different sets of neighbors. The result is 10 000 unified domain families (excluding singletons) representing functionally related proteins and recovering classical prolific domain types in high nos. The classification is useful, for example, in developing strategies for efficient database searching and for selecting targets to complete the map of all 3-D structures.
RE.CNT 22 THERE ARE 22 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 19 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:343429 CAPLUS
DN 135:119408
TI Evidence for a new Hepatitis C virus antigen encoded in an overlapping reading frame
AU Walewski, Jose L.; Keller, Toby R.; Stump, Decherd D.; Branch, Andrea D.
CS Division of Liver Diseases, Department of Medicine, Mount Sinai School of Medicine, New York, NY, 10029, USA
SO RNA (2001), 7(5), 710-721 CODEN: RNARFU; ISSN: 1355-8382
PB Cambridge University Press
DT Journal
LA English
AB Many viruses have overlapping genes and/or regions in which a nucleic acid signal is embedded in a coding sequence. To search for dual-use regions in the hepatitis C virus (HCV), we developed a facile computer-based sequence anal. method to map dual-use regions in coding sequences. Eight diverse full-length HCV RNA and polyprotein sequences were aligned and analyzed . A cluster of unusually conserved synonymous codons was found in the core-encoding region, indicating a potential overlapping open reading frame (ORF). Four peptides (A1, A2, A3, and A4) representing this alternate reading frame protein (ARFP), two others from the HCV core protein , and one from bovine serum albumin (BSA) were conjugated to BSA and used in western blots to test sera for specific antibodies from 100 chronic HCV patients, 44 healthy controls, and 60 patients with non-HCV liver disease. At a 1:20,000 diln., specific IgGs to three of the four ARFP peptides were detected in chronic HCV sera. Reactivity to either the A1 or A3 peptides (both ARFP derived) was significantly assocd. with chronic HCV infection, when compared to non-HCV liver disease serum samples (10/100 vs. 1/60; $p < 0.025$). Antibodies to A4 were not detected in any serum sample. Our western blot assays confirmed the presence of specific antibodies to a new HCV antigen encoded, at least in part, in an alternate reading frame (ARF) overlapping the core-encoding region. Because this novel HCV protein stimulates specific immune responses, it has potential value in diagnostic tests and as a component of vaccines. This protein is predicted to be highly basic and may play a role in HCV replication, pathogenesis, and carcinogenesis.
RE.CNT 38 THERE ARE 38 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 20 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:341221 CAPLUS
DN 136:49884
TI Classification and analysis of eukaryotic ABC transporters in complete eukarya genomes
AU Igarashi, Yoshinobu; Kihara, Daisuke; Kanehisa, Minoru
CS Institute for Chemical Research, Kyoto University, Kyoto, 611-0011, Japan
SO Genome Informatics Series (2000), 11(Genome Informatics 2000), 274-275 CODEN: GINSE9; ISSN: 0919-9454
PB Universal Academy Press
DT Journal
LA English
AB The eukaryotic ABC (ATP-Binding Cassette) transporters in S. cerevisiae, C. elegans and D. melanogaster, which are the three eukarya whose genomes have been sequenced completely, were classified and analyzed . The transporters were classified into orthologs and paralogs based on sequence similarity and domain structure according to the hierarchical cluster anal. Hidden Markov models (HMM) were built using individual clusters , and were used to search for similar sequences in other genomes in the KEEG/GENES database. Using the HMM search in bacteria, archaea and eukarya, a specific ATP-binding domain group was identified, whose homologs are found in S only plants and fungi. Results suggest that it is possible that N-terminal side of the sequences may have a function special to the medicine tolerance of fungi and plant cells.
RE.CNT 3 THERE ARE 3 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 21 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:331553 CAPLUS
DN 136:1289
TI AsMamDB: an alternative splice database of mammals
AU Ji, Hongkai; Zhou, Qing; Wen, Fang; Xia, Huiyu; Lu, Xin; Li, Yanda
CS Institute of Bioinformatics, Tsinghua University, Beijing, 100084, Peop. Rep. China

SO Nucleic Acids Research (2001), 29(1), 260-263 CODEN: NARHAD; ISSN: 0305-1048
PB Oxford University Press
DT Journal
LA English
AB The objective of database AsMamDB is to facilitate the systematic study of alternatively spliced genes of mammals. Version 1.0 of AsMamDB contains 1563 alternatively spliced genes of human, mouse and rat, each assocd. with a cluster of nucleotide sequences. The main information provided by AsMamDB includes gene alternative splicing patterns, gene structures, locations in chromosomes, products of genes and tissues where they express. Alternative splicing patterns are represented by multiple alignments of various gene transcripts and by graphs of their topol. structures. Gene structures are illustrated by exon, intron and various regulatory elements distributions. There are 4204 DNAs, 3977 mRNAs, 8989 CDSs and 126 931 ESTs in the current database. More than 130 000 GenBank entries are covered and 4443 MEDLINE records are linked. DNA, mRNA, exon, intron and relevant regulatory element sequences are provided in FASTA format. More information can be obtained by using the web-based multiple alignment tool Asalign and various category lists. AsMamDB can be accessed at http://166.111.30.65/ASMAMDB.html.
RE.CNT 7 THERE ARE 7 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 22 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:331541 CAPLUS
DN 136:32323
TI A rapid classification protocol for the CATH domain database to support structural genomics
AU Pearl, Frances M. G.; Martin, Nigel; Bray, James E.; Buchan, Daniel W. A.; Harrison, Andrew P.; Lee, David; Reeves, Gabrielle A.; Shepherd, Adrian J.; Sillitoe, Ian; Todd, Annabel E.; Thornton, Janet M.; Orengo, Christine A.
CS Department of Biochemistry and Molecular Biology, University College London, London, WC1E 6BT, UK
SO Nucleic Acids Research (2001), 29(1), 223-227 CODEN: NARHAD; ISSN: 0305-1048
PB Oxford University Press
DT Journal
LA English
AB In order to support the structural genomic initiatives, both by rapidly classifying newly detd. structures and by suggesting suitable targets for structure detn., we have recently developed several new protocols for classifying structures in the CATH domain database (http://www.biochem.ucl.ac.ukW bsm/cath). These aim to increase the speed of classification of new structures using fast algorithms for structure comparison (GRATH) and to improve the sensitivity in recognizing distant structural relatives by incorporating sequence information from relatives in the genomes (DomainFinder). In order to ensure the integrity of the database given the expected increase in data, the CATH Protein Family Database (CATH-PFDB), which currently includes 25 320 structural domains and a further 160 000 sequence relatives has now been installed in a relational ORACLE database. This was essential for developing more rigorous validation procedures and for allowing efficient querying of the database, particularly for genome anal. The assocd. Dictionary of Homologous Superfamilies, which provides multiple structural alignments and functional information to assist in assigning new relatives, has also been expanded recently and now includes information for 903 homologous superfamilies. In order to improve coverage of known structures, preliminary classification levels are now provided for new structures at interim stages in the classification protocol. Since a large proportion of new structures can be rapidly classified using profile-based sequence anal., this provides preliminary classification for easily recognizable homologues, which in the latest release of CATH (version 1.7) represented nearly three-quarters of the non-identical structures.
RE.CNT 14 THERE ARE 14 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 23 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:259486 CAPLUS
DN 136:97855
TI The human proteomics initiative (HPI)
AU O'Donovan, C.; Apweiler, R.; Bairoch, A.
CS Wellcome Trust Genome Campus, EMBL Outstation, The European Bioinformatics Institute, Hinxton, Cambridgeshire, CB10 1SD, UK
SO Trends in Biotechnology (2001), 19(5), 178-181 CODEN: TRBIDM; ISSN: 0167-7799
PB Elsevier Science Ltd.
DT Journal; General Review
LA English
AB A review. The availability of the human genome sequence has enabled the exploration and exploitation of the human genome and proteome to begin. Research has now focussed on the annotation of the genome and in particular of the proteome. With expert annotation extd. from the literature by biologists as the foundation, it has been possible to expand into the areas of data mining and automatic annotation. With further development and integration of pattern recognition methods and the application of alignments clustering , proteome anal. can now be provided in a meaningful way. These various approaches have been integrated to attach, ext. and combine as much relevant information as possible to the proteome. This resource should be valuable to users from both research and industry.
RE.CNT 16 THERE ARE 16 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 24 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:187436 CAPLUS
DN 135:252451
TI Genome alignment , evolution of prokaryotic genome organization, and prediction of gene function using genomic context
AU Wolf, Yuri I.; Rogozin, Igor B.; Kondrashov, Alexey S.; Koonin, Eugene V.
CS National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, 20894, USA
SO Genome Research (2001), 11(3), 356-372 CODEN: GEREFS; ISSN: 1088-9051
PB Cold Spring Harbor Laboratory Press
DT Journal
LA English
AB Gene order in prokaryotes is conserved to a much lesser extent than protein sequences. Only several operons, primarily those that code for phys. interacting proteins , are conserved in all or most of the bacterial and archaeal genomes. Nevertheless, even the limited conservation of operon organization that is obsd. can provide valuable evolutionary and functional clues through multiple genome comparisons. A program for constructing gapped local alignments of conserved gene strings in two genomes was developed. The

statistical significance of the local alignments was assessed using Monte Carlo simulations. Sets of local alignments were generated for all pairs of completely sequenced bacterial and archaeal genomes, and for each genome a template-anchored multiple alignment was constructed. In most pairwise genome comparisons, <10% of the genes in each genome belonged to conserved gene strings. When closely related pairs of species (i.e., two mycoplasmas) are excluded, the total coverage of genomes by conserved gene strings ranged from <5% for the cyanobacterium Synechocystis sp to 24% for the minimal genome of Mycoplasma genitalium, and 23% in Thermotoga maritima. The coverage of the archaeal genomes was only slightly lower than that of bacterial genomes. The majority of the conserved gene strings are known operons, with the ribosomal superoperon being the top-scoring string in most genome comparisons. However, in some of the bacterial-archaeal pairs, the superoperon is rearranged to the extent that other operons, primarily those subject to horizontal transfer, show the greatest level of conservation, such as the archaeal-type H+-ATPase operon or ABC-type transport cassettes. The level of gene order conservation among prokaryotic genomes was compared to the cooccurrence of genomes in clusters of orthologous genes (COGs) and to the conservation of protein sequences themselves. Only limited correlation was obsd. between these evolutionary variables. Gene order conservation shows a much lower variance than the cooccurrence of genomes in COGs, which indicates that intragenome homogenization via recombination occurs in evolution much faster than intergenome homogenization via horizontal gene transfer and lineage-specific gene loss. The potential of using template-anchored multiple-genome alignments for predicting functions of uncharacterized genes was quant. assessed. Functions were predicted or significantly clarified for .apprx.90 COGs (.apprx.4% of the total of 2414 analyzed COGs). The most significant predictions were obtained for the poorly characterized archaeal genomes; these include a previously uncharacterized restriction-modification system, a nuclease-helicase combination implicated in DNA repair, and the probable archaeal counterpart of the eukaryotic exosome. Multiple genome alignments are a resource for studies on operon rearrangement and disruption, which is central to our understanding of the evolution of prokaryotic genomes. Because of the rapid evolution of the gene order, the potential of genome alignment for prediction of gene functions is limited, but nevertheless, such predictions information significantly complements the results obtained through protein sequence and structure anal.
RE.CNT 45 THERE ARE 45 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 25 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:110789 CAPLUS
DN 135:206185
TI Cloning, sequence analysis and heterologous expression of the DNA adenine-(N6) methyltransferase from the human pathogen Actinobacillus actinomycetemcomitans
AU Eberhard, J.; Oza, J.; Reich, N. O.
CS Department of Operative Dentistry and Periodontology, University of Kiel, Kiel, Germany
SO FEMS Microbiology Letters (2001), 195(2), 223-229
CODEN: FMLED7; ISSN: 0378-1097
PB Elsevier Science B.V.
DT Journal
LA English
AB We cloned and sequenced the DNA adenine-N6 methyltransferase gene of the human pathogen Actinobacillus actinomycetemcomitans (M.AacDAM). Restriction digestion shows that the enzyme methylates adenine in the sequence GATC. Expression of the enzyme in a DAM- background shows in vivo activity. A PSI-BLAST search revealed that M.AacDAM is most related to M.HindIV, M.EcoDAM, M.StyDAM, and M.SmaII. The ClustalW alignment shows highly conserved regions in the enzyme characteristic for type a MTases. Phylogenetic tree anal. shows a cluster of enzymes recognizing the sequence GATC, within a branch of orphan MTases harboring M.AacDAM. The cloning and sequencing of this first methyltransferase gene described for A. actinomycetemcomitans open the path for studies on the potential regulatory impact of DNA methylation on gene regulation and virulence in this organism.
RE.CNT 22 THERE ARE 22 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 26 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:56192 CAPLUS
DN 134:291693
TI Contributions of Residue Pairing to .beta.-sheet Formation: Conservation and Covariation of Amino Acid Residue Pairs on Antiparallel .beta.-strands
AU Mandel-Gutfreund, Yael; Zaremba, Sydney M.; Gregoret, Lydia M.
CS Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA, 95064, USA
SO Journal of Molecular Biology (2001), 305(5), 1145-1159
CODEN: JMOBAK; ISSN: 0022-2836
PB Academic Press
DT Journal
LA English
AB In an effort to better understand .beta.-sheet assembly, we have investigated the evolutionary behavior of neighboring residues on adjacent antiparallel .beta.-strands. Residue pairs were classified according to solvent exposure as well as by whether their backbone NH and C[dbond]O groups are hydrogen bonded. The conservation and covariation of 19,241 pairs in 219 sequence alignments was analyzed . Buried pairs were found to be the most conserved, while stronger covariation was detected in the solvent-exposed pairs. However, residues on neighboring strands showed a degree of conservation and covariation similar to that of well-sepd. residues on the same strand, suggesting that evolutionary pressure to maintain complementarity between pairs on neighboring strands is weak. Moreover, in spite of the preference of certain amino acid pairs to occupy neighboring positions on adjacent strands, such favored pairs are neither more strongly mutually conserved nor covary more strongly than pairs of the same type in non-interacting positions. Although the .beta.-sheet pairs did not show outstanding evolutionary coupling, in many protein families significant conservation and covariation patterns were detected for some of the residue pairs. Overall, the weak evolutionary conservation and covariation of the .beta.-sheet pairs indicates that sheet structure is unlikely to be dictated by specific side-chain interactions. (c) 2001 Academic Press.
RE.CNT 54 THERE ARE 54 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 27 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2001:30527 CAPLUS
DN 135:134058
TI Molecular strategy for "serotyping" of human enteroviruses

AU Caro, Valerie; Guillot, Sophie; Delpeyroux, Francis; Crainic, Radu
CS Laboratoire d'Epidemiologie Moleculaire des Enterovirus, Institut Pasteur, Paris, 75724, Fr.
SO Journal of General Virology (2001), 82(1), 79-91 CODEN: JGVIAY; ISSN: 0022-1317
PB Society for General Microbiology
DT Journal
LA English
AB To explore further the phylogenetic relationships between human enteroviruses and to develop new diagnostic approaches, we designed a pair of generic primers in order to study a 1452 bp genomic fragment (relative to the poliovirus Mahoney genome), including the 3' end of the VP1-coding region, the 2A- and 2B-coding regions, and the 5' moiety of the 2C-coding region. Fiftynine of the 64 prototype strains and 45 field isolates of various origins, involving 21 serotypes and 6 strains untypeable by std. immunol. techniques, were successfully amplified with these primers. By detg. the nucleotide sequence of the genomic fragment encoding the C-terminal third of the VP1 capsid protein we developed a mol. typing method based on RT-PCR and sequencing. If field isolate sequences were compared to human enterovirus VP1 sequences available in databases, nucleotide identity score was, in each case, highest with the homotypic prototype (74.8 to 89.4%). Phylogenetic trees were generated from alignments of partial VP1 sequences with several phylogeny algorithms. In all cases, the new classification of enteroviruses into five identified species was confirmed and strains of the same serotype were always monophyletic. Anal. of the results confirmed that the 3' third of the VP1-coding sequence contains serotype-specific information and can be used as the basis of an effective and rapid mol. typing method. Furthermore, the amplification of such a long genomic fragment, including non-structural regions, is straightforward and could be used to investigate genome variability and to identify recombination breakpoints or specific attributes of pathogenicity.
RE.CNT 45 THERE ARE 45 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 28 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:895488 CAPLUS
DN 135:221815
TI Towards a covering set of protein family profiles
AU Heger, A.; Holm, L.
CS EMBL-EBI, Structural Genomics Group, Cambridge, CB10 1SD, UK
SO Progress in Biophysics & Molecular Biology (2000), 73(5), 321-337 CODEN: PBIMAC; ISSN: 0079-6107
PB Elsevier Science Ltd.
DT Journal; General Review
LA English
AB A review with 53 refs. Evolutionary classification leads to an economical description of the protein sequence universe because attributes of function and structure are inherited in protein families. Efficient strategies of functional and structural genomics therefore target one representative from each family. Enumerating all families and establishing family membership consistently based on sequence similarities are nontrivial computational problems. Emerging concepts and caveats of global sequence clustering are reviewed. Explicit multiple alignments coupled with neighborhood anal. lead to domain segmentation, and hierarchical unification helps to resolve conflicts and validate clusters . Eventually, every part of every sequence will be assigned to a domain family which is uniquely assocd. with a fold and a mol. function.
RE.CNT 53 THERE ARE 53 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 29 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:821389 CAPLUS
DN 133:330308
TI Analysis of the expressed genome of the lone star tick, Amblyomma americanum (Acari: Ixodidae) using an expressed sequence tag approach
AU Hill, Catherine A.; Gutierrez, Jesus A.
CS Elanco Animal Health, A Division of Eli Lilly and Company, Greenfield, IN, 46140, USA
SO Microbial & Comparative Genomics (2000), 5(2), 89-101 CODEN: MCGEFP; ISSN: 1090-6592
PB Mary Ann Liebert, Inc.
DT Journal
LA English
AB An expressed sequence tag (EST) approach was used to study the genome of two developmental stages of the lone star tick, Amblyomma americanum. The cDNA libraries were constructed from the larval and adult stages of A. americanum. In total, 1942 ESTs were sequenced (1462 adult ESTs and 480 larval ESTs) and analyzed using bioinformatic programs. Contig assembly using the CAPII program revealed 11% and 15% redundancy of sequences in the larval and adult ESTs, resp. Of the 1942 ESTs, 1738 sequences were considered quality sequences and of these, 771 or approx. 44.4% of the sequences were putatively identified based on amino acid identity using the protein Basic Local Alignment Search Tool (BLAST) algorithm. Putatively identified sequences were classified according to their predicted gene function. In total, 967 sequences, or 55.6% of the quality sequences, had limited or not protein similarity to previously identified gene products. Sequences lacking protein homol. were analyzed using an automated sequence annotation system for predicted protein characteristics such as open reading frames, signal peptides , protein motifs, and transmembrane regions. This paper describes the sequencing of the largest no. of ESTs obtained from an arachnid species to date and the subsequent detailed anal. of these sequences (GenBank Accession Nos. BF006789-BF008649).
RE.CNT 31 THERE ARE 31 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 30 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:783891 CAPLUS
DN 135:103164
TI Sequence -motif analysis of 5'-untranslated region of GB virus-C in Japanese patients
AU Kanda, Tatsuo; Yokosuka, Osamu; Kawai, Shigenobu; Imazeki, Fumio; Saisho, Hiromitsu
CS First Department of Medicine, Chiba University School of Medicine, Chiba, 260-8670, Japan
SO Journal of Gastroenterology and Hepatology (2000), 15(9), 1048-1053 CODEN: JGHEEO; ISSN: 0815-9319
PB Blackwell Science Asia Pty Ltd.
DT Journal
LA English
AB Background: GB Virus C (GBV-C) is considered to belong to the Flaviviridae; however, the structures of the N-terminal end of its putative polyprotein are not well known. The internal ribosomal entry site (IRES) at the 5'-untranslated region of GBV-C and an initiating codon at nucleotides (nt) 552-554

have been proposed. We investigated the validity of this proposal. Methods: The 5'-untranslated region of GBV-C was amplified from serum samples of 17 Japanese patients. Polymerase chain reaction-amplified products were directly sequenced and the obtained sequences were analyzed by comparing them with the IRES structure of other viruses. Results: Fifteen of the 17 (88%) GBV-C strains in our patients were classified as being Asian type. The box-A-like sequence (UUUC) and box-B-like sequence (AUCAUGG) obsd. in the IRES of picornaviruses were highly conserved in all the strains. Based on pair-wise comparisons with the multiple alignment data, overall sequence divergence for the 5'-terminus was 2.9-12%. When compared with the proposed secondary structure of the IRES model, the sequence divergences of the Asian-type GBV-C were higher at the regions of loop structures and lower at the regions of double-stranded RNA. The AUG codons, except for the one located at nt 552-554, produced truncated polyproteins or were not in-frame with the putative protein . Conclusions: Our examn. of the sequence motif of GBV-C supports the proposal that the GBV-C has common structural motifs for IRES at its 5'-untranslated region and the AUG codon at nt 552-554 may be an initiating codon.
RE.CNT 35 THERE ARE 35 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 31 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:760353 CAPLUS
DN 134:68035
TI A common molecular signature unifies the chitosanases belonging to families 46 and 80 of glycoside hydrolases
AU Tremblay, Hugo; Blanchard, Josee; Brzezinski, Ryszard
CS Centre d'Etude et de Valorisation de la Diversite Microbienne, Departement de biologie, Faculte des sciences, Universite de Sherbrooke, Sherbrooke, QC, J1K2R1, Can.
SO Canadian Journal of Microbiology (2000), 46(10), 952-955
CODEN: CJMIAZ; ISSN: 0008-4166
PB National Research Council of Canada
DT Journal
LA English
AB The 3D structure-oriented alignment of the primary sequences of fourteen chitosanases, mainly of bacterial origin and belonging to families 46 and 80 of glycoside hydrolases, resulted in the identification of the following pattern common to all these enzymes: E-[DNQ]-x(8,17)-Y- x(7)-D-x-[RD]-[GP]-x-[TS]-x(3)-[AIVFLY]-G-x(5,11)-D. This pattern is proposed as the mol. signature of the chitosanases from families 46 and 80. It includes several amino acids essential for enzyme activity and (or) stability as shown by site-directed mutagenesis studies on the chitosanase from Streptomyces sp. N174. In particular, it includes two carboxylic residues directly involved in catalysis. We suggest that there is a continuum of sequence similarity between all the analyzed chitosanases, and that all these enzymes should probably be classified in one family.
RE.CNT 16 THERE ARE 16 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 32 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:678294 CAPLUS
DN 133:346719
TI Practical limits of function prediction
AU Devos, Damien; Valencia, Alfonso
CS Protein Design Group, CNB-CSIC, Madrid, E-28049, Spain
SO Proteins: Structure, Function, and Genetics (2000), 41(1), 98-107 CODEN: PSFGEY; ISSN: 0887-3585
PB Wiley-Liss, Inc.
DT Journal
LA English
AB The widening gap between known protein sequences and their functions has led to the practice of assigning a potential function to a protein on the basis of sequence similarity to proteins whose function has been exptl. investigated. We present here a crit. view of the theor. and practical bases for this approach. The results obtained by analyzing a significant no. of true sequence similarities, derived directly from structural alignments , point to the complexity of function prediction. Different aspects of protein function, including (i) enzymic function classification , (ii) functional annotations in the form of key words, (iii) classes of cellular function, and (iv) conservation of binding sites can only be reliably transferred between similar sequences to a modest degree. The reason for this difficulty is a combination of the unavoidable database inaccuracies and the plasticity of protein function. In addn., anal. of the relationship between sequence and functional descriptions defines an empirical limit for pairwise-based functional annotations, namely, the three first digits of the six nos. used as descriptors of protein folds in the FSSP database can be predicted at an av. level as low as 7.5% sequence identity, two of the four EC digits at 15% identity, half of the SWISS-PROT key words related to protein function would require 20% identity, and the prediction of half of the residues in the binding site can be made at the 30% sequence identity level.
RE.CNT 48 THERE ARE 48 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 33 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:661815 CAPLUS
DN 134:14471
TI cDNA cloning and sequence of European sea bass (Dicentrarchus labrax) somatolactin
AU Company, R.; Calduch-Giner, J. A.; Mingarro, M.; Perez-Sanchez, J.
CS Instituto de Acuicultura de Torre de la Sal (CSIC), Ribera de Cabanes, Castellon, 12595, Spain
SO Comparative Biochemistry and Physiology, Part B: Biochemistry & Molecular Biology (2000), 127B(2), 183-192
CODEN: CBPBB8; ISSN: 0305-0491
PB Elsevier Science Inc.
DT Journal
LA English
AB Three cDNA clones encoding for European sea bass somatolactin (SL) were obtained by RT-PCR and 3' RACE of RNA of pituitary origin. Clone 1 was 582 bp in length, and included a part of the signal peptide and the 5' end of the mature protein . Clone 2 (1075 bp) included a fragment of the coding sequence and the 3' untranslated region, which was 888 bp in length and contained two putative polyadenylation signals (AATAAA) at 12-17, and 202-207 nucleotides upstream of the poly (A) tail. Clone 3 was 624 bp in length and its nucleotide sequence encoding for the entire mature protein confirmed the sequence already detd. from the first two clones. The size of SL mRNA transcripts was estd. by Northern blot anal. and a single band of approx. 1.6 kb was obsd. with pituitary RNAs. No band was found with RNAs of brain and liver origin. Alignment of the deduced amino acid sequence revealed that European sea bass SL shared 90-84% identity with perciform, pleuronectiform and scorpaeniform fish SLs, and 77-57% with other SLs of more distant fish orders, with a strict conservation of Cys residues and the N-glycosylation site (Asn-Lys-Thr) at 121-123 amino acid positions. The

reconstruction of the phylogenetic tree based on SL nucleotide sequences , and analyzed by max. likelihood distances, showed the same clustering as the present hierarchy of fish. When comparisons were made among SL, prolactin and growth hormone of European sea bass, the overall amino identity was relatively low (22-23%). However, a high degree of amino acid homol. was found at the C-terminus, which contains three of the four Cys residues strictly conserved in all the members of GH/PRL family.
RE.CNT 32 THERE ARE 32 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 34 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:545507 CAPLUS
DN 133:249970
TI Sequence and phylogenetic analysis of squid myosin-V: A vesicle motor in nerve cells
AU Molyneaux, Bradley J.; Mulcahey, Mary K.; Stafford, Phillip; Langford, George M.
CS Department of Biological Sciences, Dartmouth College, Hanover, NH, USA
SO Cell Motility and the Cytoskeleton (2000), 46(2), 108-115
CODEN: CMCYEO; ISSN: 0886-1544
PB Wiley-Liss, Inc.
DT Journal
LA English
AB Expts. were performed to clone and sequence the cDNA for squid brain myosin V. Five proteolytic fragments of purified squid brain myosin V were analyzed by direct protein sequencing. Based on this sequence information, degenerate primers were constructed and used to isolate cDNA clones by PCR. Five clones, representing overlapping segments of the gene, were sequenced. The sequence data and the previous biochem. characterization of the mol. support the classification of this vesicle-assocd. myosin as a member of the class V myosins. Motif anal. of the head, neck, and tail domains revealed that squid MyoV has consensus sequences for all the motifs found in vertebrate members of the myosin V family of motor proteins . A phylogenetic tree was constructed from a sequence alignment by the neighbor-joining method, using Megalign; the resulting phylogenetic tree showed that squid MyoV is more closely related to vertebrate MyoV (mouse dil., chicken dil., rat myr6, and human myo5a) than Drosophila and yeast (myo2, and myo4) myosins V. These new data on the phylogenetic relationships of squid myosin V to vertebrate myosin V strengthens the argument that myosin V functions as a vesicle motor in vertebrate neurons.
RE.CNT 16 THERE ARE 16 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 35 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:539646 CAPLUS
DN 134:247662
TI Determination of human immunodeficiency virus type 1 subtypes in Taiwan by vpu gene analysis
AU Lee, Chun-Nan; Wang, Wei-Kung; Fan, Wen-Sheng; Twu, Shing-Jer; Chen, Shou-Chien; Sheng, Ming-Ching; Chen, Mao-Yuan
CS School and Graduate Institute of Medical Technology, National Taiwan University, Taipei, 100, Taiwan
SO Journal of Clinical Microbiology (2000), 38(7), 2468-2474
CODEN: JCMIDW; ISSN: 0095-1137
PB American Society for Microbiology
DT Journal
LA English
AB The genetic diversity of human immunodeficiency virus (HIV) type 1 (HIV-1) has been characterized mainly by anal. of the env and gag genes. Information on the vpu genes in the HIV sequence database is very limited. In the present study, the nucleotide sequences of the vpu genes were analyzed , and the genetic subtypes detd. by anal. of the vpu gene were compared with those previously detd. by anal. of the gag and env genes. The vpu genes were amplified by nested PCR of proviral DNA extd. from 363 HIV-1-infected individuals and were sequenced directly by use of the PCR products. HIV-1 subtypes were detd. by sequence alignment and phylogenetic anal. with ref. strains. The strains in all except one of the samples analyzed could be classified as subtype A, B, C, E, or G. The vpu subtype of one strain could not be detd. Of the strains analyzed, genetic subtypes of 247 (68.0%) were also detd. by anal. of the env or gag gene. The genetic subtypes detd. by vpu gene anal. were, in general, consistent with those detd. by gag and/or env gene anal. except for those for two AG recombinant strains. All the strains that clustered with a Thailand subtype E strain in the vpu phylogenetic analyses were subtype E by env gene anal. and subtype A by gag gene anal. In summary, our genetic typing revealed that subtype B strains, which constituted 73.8% of all strains analyzed, were most prevalent in Taiwan. While subtype E strains constituted about one-quarter of the viruses, they were prevalent at a higher proportion in the group infected by heterosexual transmission. Genetic anal. of vpu may provide an alternate method for detn. of HIV-1 subtypes for most of the strains, excluding those in which intersubtype recombination has occurred.
RE.CNT 26 THERE ARE 26 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 36 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:398091 CAPLUS
DN 133:116608
TI Assignment of enzyme substrate specificity by principal component analysis of aligned protein sequences: an experimental test using DNA glycosylase homologs
AU Gogos, Arhonda; Jantz, Derek; Senturker, Sema; Richardson, Delwood; Dizdaroglu, Miral; Clarke, Neil D.
CS Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, Baltimore, MD, 21205, USA
SO Proteins: Structure, Function, and Genetics (2000), 40(1), 98-105 CODEN: PSFGEY; ISSN: 0887-3585
PB Wiley-Liss, Inc.
DT Journal
LA English
AB We have studied the relationship between amino acid sequence and substrate specificity in a DNA glycosylase family by characterizing exptl. the specificity of four new members of the family. We show that principal component anal. (PCA) of the sequence family correctly predicts the substrate specificity of one of the novel homologs even though conventional sequence anal. methods fail to group this homolog with other sequences of the same specificity. PCA also suggested, correctly, that another homolog characterized previously differs in its specificity from those sequences with which it clusters by conventional criteria. These results suggest that principal component anal. of sequence families can be a useful tool in annotating genome sequences when there is ambiguity concerning which subfamily a new homolog belongs to.
RE.CNT 44 THERE ARE 44 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 37 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:359646 CAPLUS
DN 133:277019
TI Sequencing and analysis of the Methylococcus capsulatus (bath) soluble methane monooxygenase genes
AU Coufal, David E.; Blazyk, Jessica L.; Whittington, Douglas A.; Wu, Wayne W.; Rosenzweig, Amy C.; Lippard, Stephen J.
CS Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
SO European Journal of Biochemistry (2000), 267(8), 2174-2185 CODEN: EJBCAI; ISSN: 0014-2956
PB Blackwell Science Ltd.
DT Journal
LA English
AB The sol. methane monooxygenase (sMMO) hydroxylase is a prototypical member of the class of proteins with non-heme carboxylate-bridged diiron sites. The sMMO subclass of enzyme systems has several distinguishing characteristics, including the ability to catalyze hydroxylation or epoxidn. chem., a multisubunit hydroxylase contg. diiron centers in its .alpha. subunits, and the requirement of a coupling protein for optimal activity. Sequence homol. alignment of known members of the sMMO family was performed in an effort to identify protein regions giving rise to these unique features. DNA sequencing of the Methylococcus capsulatus (Bath) sMMO genes confirmed previously identified sequencing errors and cor. two addnl. errors, each of which was confirmed by at least one independent method. Alignments of homologous proteins from sMMO, phenol hydroxylase, toluene 2-, 3-, and 4-monooxygenases, and alkene monooxygenase systems revealed an interesting set of absolutely conserved amino-acid residues, including previously unidentified residues located outside the diiron active site of the hydroxylase. By mapping these residues on to the M. capsulatus (Bath) sMMO hydroxylase crystal structure, functional and structural roles were proposed for the conserved regions. Anal. of the active site showed a highly conserved hydrogen-bonding network on one side of the diiron cluster but little homol. on the opposite side, where substrates are presumed to bind. It is suggested that conserved residues on the hydroxylase surface may be important for protein - protein interactions with the reductase and coupling ancillary proteins and/or serve as part of an electron-transfer pathway. A possible way by which binding of the coupling protein at the surface of the hydroxylase might transfer information to the diiron active site at the interior is proposed.
RE.CNT 74 THERE ARE 74 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 38 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:342202 CAPLUS
DN 133:13417
TI A chemically synthesized artificial promotor for high level expression of transgenes and a method for its synthesis
IN Tuli, Rakesh; Sawant, Samir Vishwanat; Singh, Pradhyumna Kumar; Gupta, Shiv Kumar
PA India
SO Jpn. Kokai Tokkyo Koho, 24 pp. CODEN: JKXXAF
DT Patent
LA Japanese
FAN.CNT 1 PATENT NO. KIND DATE APPLICATION NO. DATE
---------- ---- -------- ---------- --------
PI JP 2000139477 A2 20000523 JP 1999-119227 19990427
PRAI IN 1998-3322 A 19981109
AB The invention relates to a chem. synthesized artificial promoter comprising a DNA sequence designed for the level and pattern of target gene expression, by strategically putting together several signature sequences identified by sequence alignment and statistical anal. of a large database constructed for this purpose. Also claimed are a method of synthesizing such a promoter and a method for testing the high level gene expression by the promoter compared to the naturally occurring CaMV 35S promoter using a GUS assay. The design includes classifying genes into high expression and low expression categories, and identifying a conserved domain among high expression genes with regard to some important elements. An artificial promoter designed and synthesized by the method induced 3-4 fold higher expression of uidA gene in tobacco protoplast and 16 fold higher expression in tobacco leaf, as well as in other plant systems. The promoter showed a high activity in tobacco leaf, stem, and root, particularly in root.

L9 ANSWER 39 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:306815 CAPLUS
DN 133:115665
TI Genome annotation assessment in Drosophila melanogaster
AU Reese, Martin G.; Hartzell, George; Harris, Nomi L.; Ohler, Uwe; Abril, Josep F.; Lewis, Suzanna E.
CS Berkeley Drosophila Genome Project, Department of Molecular and Cell Biology, University of California, Berkeley, CA, 94720-3200, USA
SO Genome Research (2000), 10(4), 483-501 CODEN: GEREFS; ISSN: 1088-9051
PB Cold Spring Harbor Laboratory Press
DT Journal
LA English
AB Computational methods for automated genome annotation are crit. to our community's ability to make full use of the large vol. of genomic sequences being generated and released. To explore the accuracy of these automated feature prediction tools in the genomes of higher organisms, we evaluated their performance on a large, well-characterized sequence contig from the Adh region of Drosophila melanogaster. This expt., known as the Genome Annotation Assessment Project (GASP), was launched in May 1999. Twelve groups, applying state-of-the-art tools, contributed predictions for features including gene structure, protein homologies, promoter sites, and repeat elements. We evaluated these predictions using two stds., one based on previously unreleased high-quality full-length cDNA sequences and a second based on the set of annotations generated as part of an in-depth study of the region by a group of Drosophila experts. Although these std. sets only approx. the unknown distribution of features in this region, we believe that when taken in context the results of an evaluation based on them are meaningful. The results were presented as a tutorial at the conference on Intelligent Systems in Mol. Biol. (ISMB-99) in August 1999. Over 95% of the coding nucleotides in the region were correctly identified by the majority of the gene finders, and the correct intron/exon structures were predicted for >40% of the genes. Homol.-based annotation techniques recognized and assocd. functions with almost half of the genes in the region; the remainder were only identified by the ab initio techniques. This expt. also presents the first assessment of promoter prediction techniques for a significant no. of genes in a large contiguous region. We discovered that the promoter predictors' high false-pos. rates make their predictions difficult to use. Integrating gene finding and cDNA/EST alignments with promoter predictions decreases the no. of false-pos. classifications but discovers less than one-third of the promoters in the region. We believe that by establishing stds. for evaluating genomic annotations and by

assessing the performance of existing automated genome annotation tools, this expt. establishes a baseline that contributes to the value of ongoing large-scale annotation projects and should guide further research in genome informatics.
RE.CNT 60 THERE ARE 60 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 40 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:250124 CAPLUS
DN 133:70370
TI Sensitive sequence comparison as protein function predictor
AU Pawlowski, K.; Jaroszewski, L.; Rychlewski, L.; Godzik, A.
CS The Burnham Institute, La Jolla, CA, 92037, USA
SO Pacific Symposium on Biocomputing 2000, Honolulu, Jan 4-9, 2000 (2000), 42-53. Editor(s): Altman, Russ B. Publisher: World Scientific Publishing Co. Pte. Ltd., Singapore, Singapore.
CODEN: 68UQA8
DT Conference
LA English
AB Protein function assignments based on postulated homol. as recognized by high sequence similarity are used routinely in genome anal. Improvements in sensitivity of sequence comparison algorithms got to the point that proteins with previously undetectable sequence similarity, such as for instance 10-15% of identical residues, sometimes can be classified as similar. What is the relation between such proteins . Is it possible that they are homologous. What is the practical significance of detecting such similarities. A simplified anal. of the relation between sequence similarity and function similarity is presented here for the well-characterized proteins from the E. coli genome. Using a simple measure of functional similarity based on E.C. classification of enzymes, it is shown that it correlates well with sequence similarity measured by statistical significance of the alignment score. Proteins , similar by this std., even in cases of low sequence identity, have a much larger chance of having similar function than the randomly chosen protein pairs. Interesting exceptions to these rules are discussed.
RE.CNT 29 THERE ARE 29 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 41 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:154089 CAPLUS
DN 132:318499
TI Characterization of a ubiquitous expressed gene family encoding polygalacturonase in Arabidopsis thaliana
AU Torki, Moez; Mandaron, Paul; Mache, Regis; Falconet, Denis
CS Laboratoire de Genetique Moleculaire des Plantes, UMR 5575, CNRS and Universite Joseph Fourier, Grenoble, F-38041, Fr.
SO Gene (2000), 242(1-2), 427-436 CODEN: GENED6; ISSN: 0378-1119
PB Elsevier Science B.V.
DT Journal
LA English
AB Pectin, as one of the major components of plant cell wall, has been implicated in many developmental processes occurring during plant growth. Among the different enzymes known to participate in the pectin structure modifications, polygalacturonase (PG) activity has been shown to be assocd. with fruit ripening, organ abscission and pollen grain development. Until now, sequence analyses of the deduced polypeptides of the plant PG genes allowed their grouping into three clades corresponding to genes involved in one of these three activities. In this study, we report the sequence of three genomic clones encoding PG in Arabidopsis thaliana. These genes, together with 16 other genes present in the databases form a large gene family, ubiquitously expressed, present on the five chromosomes with at least two gene clusters on chromosomes II and V, resp. Phylogenetic analyses suggest that the A. thaliana gene family contains five classes of genes, with three of them corresponding to the previously defined clades. Comparison of positions and nos. of introns among the A. thaliana genes reveals structural conservation between genes belonging to the same class. The pattern of intron losses that could have given rise to the PG gene family is consistent with a mechanism of intron loss by replacement of an ancestral intron-contg. gene with a reverse-transcribed DNA copy of a spliced mRNA. Following this event of intron loss, the acquisition of introns in novel positions is consistent with a mechanism of intron gain at proto-splice sites.
RE.CNT 28 THERE ARE 28 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 42 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:82735 CAPLUS
DN 132:247033
TI Genetic diversity and the absence of regional differences of Borrelia garinii as demonstrated by ospA and ospB gene sequence analysis
AU Yabuki, Mihe; Nakao, Minoru; Fukunaga, Masahito
CS Laboratory of Molecular Microbiology, Faculty of Pharmacy and Pharmaceutical Sciences, Fukuyama University, Hiroshima, 729-0292, Japan
SO Microbiology and Immunology (1999), 43(12), 1097-1102
CODEN: MIIMDV; ISSN: 0385-5600
PB Center for Academic Publications Japan
DT Journal
LA English
AB Unfed adult Ixodes persulcatus ticks were collected from four locations of Nagano and Hokkaido in Japan. Infected Borrelia garinii were investigated by PCR-RFLP of the ospA and ospB gene sequences. The primer set amplified an approx. 1.6-kb DNA fragment (0.7-kb in some strains), and BsrI, BstYI, or NlaIII digestion of the product resulted in six distinctively different PCR-RFLP groups and two independent borrelial strains. The representatives in each PCR-RFLP group and individuals from the borrelial strains were sequenced, and their deduced amino acid sequences were aligned . A neighbor-joining phylogenetic anal. showed that the B. garinii OspA or OspB sequences were each divided into three major clusters including isolates from both the Nagano and Hokkaido locations. There was no local difference in OspA/B sequences between Nagano and Hokkaido. The osp gene of Borrelia burgdorferi sensu lato is highly heterogeneous, and this was also confirmed by our sequence anal. Some strains of the different PCR-RFLP groups had closely related OspA sequences, while the OspB sequences of these strains were quite different. These findings suggested intraspecies gene exchange and recombination events between the two genes in B. garinii.
RE.CNT 20 THERE ARE 20 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 43 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:74574 CAPLUS
DN 133:100161

TI Typing of Candida glabrata in clinical isolates by comparative sequence analysis of the cytochrome c oxidase subunit 2 gene distinguishes two clusters of strains associated with geographical sequence polymorphisms
AU Sanson, Gerdine F. O.; Briones, Marcelo R. S.
CS Disciplina de Microbiologia, Universidade Federal de Sao Paulo, Sao Paulo, Brazil
SO Journal of Clinical Microbiology (2000), 38(1), 227-235
CODEN: JCMIDW; ISSN: 0095-1137
PB American Society for Microbiology
DT Journal
LA English
AB The authors tested whether comparative sequence anal. of the mitochondrion-encoded cytochrome c oxidase subunit 2 gene (COX2) could be used to distinguish intraspecific variants of Candida glabrata. Mitochondrial genes are suitable for investigation of close phylogenetic relationships because they evolve much faster than nuclear genes, which in general exhibit very limited intraspecific variation. For this survey the authors used 11 clin. isolates of C. glabrata from three different geog. locations in Brazil, 10 isolates from one location in the United States, 1 American Type Culture Collection strain as an internal control, and the published sequence of strain CBS 138. The complete coding region of COX2 was amplified from total cellular DNA, and both strands were sequenced twice for each strain. These sequences were aligned with published sequences from other fungi, and the nos. of substitutions and phylogenetic relationships were detd. Typing of these strains was done by using 17 substitutions, with 8 being nonsynonymous and 9 being synonymous. Also, cDNAs made from purified mitochondrial polyadenylated RNA were sequenced to confirm that the sequences correspond to the expressed copies and not nuclear pseudogenes and that a frameshift mutation exists in the 3' end of the coding region (position 673) relative to the Saccharomyces cerevisiae sequence and the previously published C. glabrata sequence. The authors estd. the av. evolutionary rate of COX2 to be 11.4% sequence divergence/108 years and that phylogenetic relationships of yeasts based on these sequences are consistent with rRNA sequence data. The anal. of COX2 sequences enables typing of C. glabrata strains based on 13 haplotypes and suggests that positions 51 and 519 indicate a geog. polymorphism that discriminates strains isolated in the United States and strains isolated in Brazil. This provides for the first time a means of typing of Candida strains that cause infections by use of direct sequence comparisons and the assocd. divergence ests.
RE.CNT 34 THERE ARE 34 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 44 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 2000:11861 CAPLUS
DN 132:162778
TI Structural determinants in domain II of human glutathione transferase M2-2 govern the characteristic activities with aminochrome, 2-cyano-1,3-dimethyl-1-nitrosoguanidine, and 1,2-dichloro-4-nitrobenzene
AU Hansson, Lars O.; Bolton-Grob, Robyn; Widersten, Mikael; Mannervik, Bengt
CS Department of Biochemistry, Biomedical Center, Uppsala University, Uppsala, S-751 23, Swed.
SO Protein Science (1999), 8(12), 2742-2750 CODEN: PRCIEI; ISSN: 0961-8368
PB Cambridge University Press
DT Journal
LA English
AB Two human Mu class glutathione transferases, hGST M1-1 and hGST M2-2, with high sequence identity (84%) exhibit a 100-fold difference in activities with the substrates aminochrome, 2-cyano-1,3-dimethyl-1-nitrosoguanidine (cyanoDMNG), and 1,2-dichloro-4-nitrobenzene (DCNB), with hGST M2-2 being more efficient. A sequence alignment with the rat Mu class GST M3-3, an enzyme also showing high activities with aminochrome and DCNB, demonstrated an identical structural cluster of residues 164-168 in the .alpha.6-helixes of rGST M3-3 and hGST M2-2, a motif unique among known sequences of human, rat, and mouse Mu class GSTs. A putative electrostatic network Arg107-Asp161-Arg165-Glu164(-Gln167) was identified based on the published three-dimensional structure of hGST M2-2. Corresponding variant residues of hGST M1-1 (Leu165, Asp164, and Arg167) as well as the active site residue Ser209 were targeted for point mutations, introducing hGST M2-2 residues to the framework of hGST M1-1, to improve the activities with substrates characteristic of hGST M2-2. In addn., chimeric enzymes composed of hGST M1-1 and hGST M2-2 sequences were analyzed . The activity with 1-chloro-2,4-dinitrobenzene (CDNB) was retained in all mutant enzymes, proving that they were catalytically competent, but none of the point mutations improved the activities with hGST M2-2 characteristic substrates. The chimeric enzymes showed that the structural determinants of these activities reside in domain II and that residue Arg165 in hGST M2-2 appears to be important for the reactions with cyanoDMNG and DCNB. A mutant, which contained all the hGST M2-2 residues of the putative electrostatic network, was still lacking one order of magnitude of the activities with the characteristic substrates of wild-type hGST M2-2. It was concluded that a limited set of point mutations is not sufficient, but that indirect secondary structural affects also contribute to the hGST M2-2 characteristic activities with aminochrome, cyanoDMNG, and DCNB.
RE.CNT 42 THERE ARE 42 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 45 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:799248 CAPLUS
DN 132:89754
TI Modular assembly of voltage-gated channel proteins : a sequence analysis and phylogenetic study
AU Nelson, Richard D.; Kuan, Gary; Saier, Milton H., Jr.; Montal, Mauricio
CS Department of Biology, University of California at San Diego, La Jolla, CA, 92093, USA
SO Journal of Molecular Microbiology and Biotechnology (1999), 1(2), 281-287 CODEN: JMMBFF; ISSN: 1464-1801
PB Horizon Scientific Press
DT Journal
LA English
AB Voltage-sensitive cation-selective ion channels of the voltage-gated ion channel (VGC) superfamily were examd. by a combination of sequence alignment and phylogenetic tree construction procedures. Segments of the .alpha.-subunits of K+-selective channels homologous to the structurally elucidated KcsA channel of Streptomyces lividans were multiply aligned , and this alignment provided the database for computer-assisted structural analyses and phylogenetic tree construction. Similar analyses were conducted with the four homologous repeats of the .alpha.-subunits from representative Ca2+- and Na+-selective channels, as well as with the ensemble of K+, Ca2+ and Na+ channels. In both the single subunit of the K+ channels and the individual

repeats of the Ca2+ and Na+ channels, the analyses suggest the occurrence of at least two tandemly arranged modules corresponding to the predicted voltage-sensor domain and the pore domain. The phylogenetic analyses reveal strict clustering of segments according to cation-selectivity and repeat unit. The authors surmise that the pore module of the prokaryotic K+ channel was the primordial polypeptide upon which other modules were superimposed during evolution in order to generate phenotypic diversity. These observations may prove applicable to all members of the VGC family yet to be discovered throughout the prokaryotic and eukaryotic kingdoms.
RE.CNT 60 THERE ARE 60 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 46 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:691808 CAPLUS
DN 132:10468
TI Dictionary building via unsupervised hierarchical motif discovery in the sequence space of natural proteins
AU Rigoutsos, Isidore; Floratos, Aris; Ouzounis, Christos; Gao, Yuan; Parida, Laxmi
CS Bioinformatics and Pattern Discovery Group, Computational Biology Center, IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY, 10598, USA
SO Proteins: Structure, Function, and Genetics (1999), 37(2), 264-277 CODEN: PSFGEY; ISSN: 0887-3585
PB Wiley-Liss, Inc.
DT Journal
LA English
AB Using TEIRESIAS, a pattern discovery method that identifies all motifs present in any given set of protein sequences without requiring alignment or explicit enumeration of the soln. space, we have explored the GenPept sequence database and built a dictionary of all sequence patterns with two or more instances. The entries of this dictionary, henceforth named seqlets, cover 98.12% of all amino acid positions in the input database and in essence provide a comprehensive finite set of descriptors for protein sequence space. As such, seqlets can be effectively used to describe almost every naturally occurring protein . In fact, seqlets can be thought of as building blocks of protein mols. that are a necessary (but not sufficient) condition for function or family equivalence memberships. Thus, seqlets can either define conserved family signatures or cut across mol. families and previously undetected sequence signals deriving from functional convergence. Moreover, we show that seqlets also can capture structurally conserved motifs. The availability of a dictionary of seqlets that has been derived in such an unsupervised, hierarchical manner is generating new opportunities for addressing problems that range from reliable classification and the correlation of sequence fragments with functional categories to faster and sensitive engines for homol. searches, evolutionary studies, and protein structure prediction.
RE.CNT 53 THERE ARE 53 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 47 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:533842 CAPLUS
DN 131:296550
TI The immunoglobulin fold family: sequence analysis and 3D structure comparisons
AU Halaby, D. M.; Poupon, A.; Mornon, J.-P.
CS Systemes Moleculaires et Biologie Structurale, LMCP, CNRS UMR C7590 Universites Pierre et Marie Curie (P6) et Denis Diderot (P7), Paris, 75252, Fr.
SO Protein Engineering (1999), 12(7), 563-571 CODEN: PRENE9; ISSN: 0269-2139
PB Oxford University Press
DT Journal; General Review
LA English
AB A review with .apprx.40 refs. Fifty-two 3D structures of Ig-like domains covering the Ig fold family (IgFF) were compared and classified according to the conservation of their secondary structures. Members of the IgFF are distantly related proteins or evolutionarily unrelated proteins with a similar fold, the Ig fold. In this paper, a multiple structural alignment of the conserved common core is described and the correlation between corresponding sequences is discussed. While the members of the IgFF exhibit wide heterogeneity in terms of tissue and species distribution or functional implications, the 3D structures of these domains are far more conserved than their sequences. We define topol. equiv. residues in the Ig-like domains, describe the hydrophobic common cores and discuss the presence of addnl. strands. The disulfide bridges, not necessary for the stability of the Ig fold, may have an effect on the compactness of the domains. Based upon sequence and structure anal., we propose the introduction of two new subtypes (C3 and C4) to the previous classifications , in addn. to a new global structural classification . The very low mean sequence identity between subgroups of the IgFF suggests the occurrence of both divergent and convergent evolutionary processes, explaining the wide diversity of the superfamily. Finally, this review suggest that hydrophobic residues constituting the common hydrophobic cores are important clues to explain how highly divergent sequences can adopt a similar fold.
RE.CNT 39 THERE ARE 39 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 48 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:409770 CAPLUS
DN 131:225174
TI Comparative Sequence Analysis of the Complete Human Sarcomeric Myosin Heavy Chain Family: Implications for Functional Diversity
AU Weiss, Allison; Schiaffino, Stefano; Leinwand, Leslie A.
CS Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY, 10461, USA
SO Journal of Molecular Biology (1999), 290(1), 61-75 CODEN: JMOBAK; ISSN: 0022-2836
PB Academic Press
DT Journal
LA English
AB The conventional myosin motor proteins that drive mammalian skeletal and cardiac muscle contraction include eight sarcomeric myosin heavy chain (MyHC) isoforms. Six skeletal MyHCs are encoded by genes found in tightly linked clusters on human and mouse chromosomes 17 and 11, resp. The full coding regions of only two out of six mammalian skeletal MyHCs had been sequenced prior to this work. In an effort to assess the extent of sequence diversity within the human MyHC family we present new full-length coding sequences corresponding to four addnl. human genes: MyHC-IIb, MyHC-extraocular, MyHC-IIa and MyHC-IIx/d. This represents the first opportunity to compare the full coding sequences of all eight sarcomeric MyHC isoforms within a vertebrate organism. Sequence variability has been analyzed in the context of available structure/function data with an

emphasis on potential functional diversity within the family. Results indicate that functional diversity among MyHCs is likely to be accomplished by having small pockets of sequence diversity in an otherwise highly conserved mol. (c) 1999 Academic Press.
RE.CNT 88 THERE ARE 88 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 49 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:395823 CAPLUS
DN 131:226562
TI Molecular genetic characterization of two insular Asian cat species, Bornean bay cat and Iriomote cat
AU Johnson, Warren E.; Shiny Ashiki, Fumiharu; Raymond, Marilyn Menotti; Driscoll, Carlos; Leh, Charles; Sunquist, Mel; Johnston, Leslie; Bush, Mitchel; Wildt, David; Yuhki, Naoya; O'Brien, Stephen J.
CS Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD, 21702-1201, USA
SO Evolutionary Theory and Processes: Modern Perspectives (1999), 223-248. Editor(s): Wasser, Solomon P. Publisher: Kluwer, Dordrecht, Neth. CODEN: 67UDAU
DT Conference
LA English
AB Mol. genetic data were used to characterize the genetic distinctiveness of Bornean bay cat (Pardofelis badia) and Iriomote cat (Prionailurus bengalensis iriomotensis), small cat species restricted to sep. Asian islands. Sequence variation in two mitochondrial genes, NADH dehydrogenase subunit 5 (NADH-5) and ATPase-8 (ATP-8) was used to examine the phylogenetic relationship between a recently discovered Bornean bay cat specimen and the original type specimen (collected in 1855) relative to other Southeast Asian felids. DNA and amino acid sequence analyses affirmed that both bay cat specimens derived from the same phylogenetic lineage and that Bornean bay cat shared a monophyletic common ancestor with Asian golden cat (Profelis temmincki) estd. at 4.9-5.3 million years ago, well before the geol. sepn. of Borneo from mainland Asia which occurred in the late Pleistocene, estd. as 10,000-20,000 yr ago. The phylogenetic distinctiveness of the Iriomote cat (Prionailurus iriomotensis or P. bengalensis iriomotensis, n=5) from two leopard cat subspecies (P. b. euptilurus, n=5 and P. b. bengalensis, n=13) was examd. based upon the DNA sequence variation of four mitochondrial genes, NADH-5, ATP-8, 16S rRNA, and Cytochrome b and based upon allele variation at 18 nuclear microsatellite loci. The available sample of Iriomote cats displayed a remarkable redn. in overall genetic diversity from diversity in both mtDNA and microsatellite variation compared to other felids. Nonetheless, the Iriomote cat genes clearly aligned them with, but distinct from, other subspecies of leopard cat (P. b. euptilurus and P. b. bengalensis) affirming their taxonomic classification as P. b. iriomotensis, subspecies. The contrasting patterns of the genetic variation of Bornean bay cat and Iriomote cat likely reflect different natural histories for these two island cat taxa.
RE.CNT 32 THERE ARE 32 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 50 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:220581 CAPLUS
DN 130:334987
TI Sequence classification of water channels and related proteins in view of functional predictions
AU Tallur, B.; Nicolas, J.; Froger, A.; Thomas, D.; Delamarche, C.
CS IRISA, Rennes, F-35042, Fr.
SO Theoretical Chemistry Accounts (1999), 101(1-3), 77-81 CODEN: TCACFW; ISSN: 1432-881X
PB Springer-Verlag
DT Journal
LA English
AB The authors have worked with a classification method based upon a notion of probabilistic similarity or "likelihood of similarity" between aligned sequences. One important parameter, among others, affecting the sequence similarities and, hence, the classification results, is the amino acid similarity matrix. The authors present a method for choosing the most adapted matrix to classify protein sequences. This method was applied to the transmembrane channels of the major intrinsic protein (MIP) family. At present, two functional subgroups are well characterized in this family: (1) specific water transport by the aquaporins and (2) small neutral solutes transport. The usefulness of the classification method in the prediction of sequence segments important for substrate selectivity was shown. The authors show that this method can also be used to predict the function of undetd. MIP proteins . The method could be applied to other protein families as well.
RE.CNT 19 THERE ARE 19 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 51 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:180263 CAPLUS
DN 131:83677
TI Cloning and nucleotide sequence analysis of psbD/C operon from chloroplasts of Populus deltoides
AU Reddy, M. S. Srinivasa; Trivedi, Prabodh K.; Tuli, Rakesh; Sane, Prafullachandra V.
CS Centre for Plant Molecular Biology, National Botanical Research Institute, Lucknow, 226 001, India
SO Journal of Genetics (1998), 77(2 & 3), 77-83 CODEN: JOGNAU; ISSN: 0022-1333
PB Indian Academy of Sciences
DT Journal
LA English
AB We report the cloning and nucleotide sequence anal. of psbD/C operon from a dicotyledonous tree species, Populus deltoides (poplar). The coding regions of psbD and psbC and deduced amino acid sequences show very high homol. with those from other higher plants. In pairwise alignment of the gene sequences, P. deltoides clustered with dicotyledonous annuals rather than with Pinus, the only other tree whose psbD/C nucleotide sequence is available. Comparison of several reported sequences showed that synonymous substitutions were distributed in both psbD and psbC uniformly, throughout the length of the genes. The frequency of nonsynonymous substitutions located in the amino-terminal end of psbD was distinctly higher, suggesting a lower degree of structural constraints in this region of the encoded D2 protein . The arrangement of reading frames and Northern anal. suggest that organization and expression of psbD/C operon in P. deltoides is similar to that in other higher plants.
RE.CNT 30 THERE ARE 30 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 52 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:175422 CAPLUS
DN 131:129514

TI Parallel protein information analysis (PAPIA) system running on a 64-node PC cluster
AU Akiyama, Yutaka; Onizuka, Kentaro; Noguchi, Tamotsu; Ando, Makoto
CS Parallel Application Laboratory, Real World Computing Partnership (RWCP), Tsukuba, 305-0032, Japan
SO Genome Informatics Series (1998), 9, 131-140 CODEN: GINSE9; ISSN: 0919-9454
PB Universal Academy Press
DT Journal
LA English
AB Protein information anal. is used widely as a key technol. in drug design, macromol. engineering and understanding genome sequences. Because a vast no. of computations are needed, further speed-up for protein information anal. is very much in demand. The PAPIA (PArallel Protein Information Anal.) system was implemented on the RWC PC cluster IIa which is composed of 65 Pentium Pro 200 MHz microprocessors. The PAPIA system performs fast parallel processing for typical computations in protein anal., such as structure similarity search, sequence homol. search and multiple sequence alignment , nearly 60 times faster than a single processor.
RE.CNT 16 THERE ARE 16 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 53 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:150512 CAPLUS
DN 130:333518
TI Sequence analysis of rabbit hemorrhagic disease virus (RHDV) in Australia: alterations after its release
AU Asgari, S.; Hardy, J. R. E.; Cooke, B. D.
CS Department of Crop Protection, The University of Adelaide, Glen Osmond, Australia
SO Archives of Virology (1999), 144(1), 135-145 CODEN: ARVIDF; ISSN: 0304-8608
PB Springer-Verlag Wien
DT Journal
LA English
AB Liver samples from rabbits killed by RHDV, collected from five States in Australia in 1996 and 1997 were analyzed by RT-PCR. A 398 bp fragment of the capsid protein (VP60) gene was amplified by PCR and directly sequenced. The alignment of the nucleotide and amino acid sequences and their comparison with the original strain of the virus released in Australia indicated genetic changes after two years have been small with 98.2% to 100% identity. The constructed phylogenetic tree suggests slight differences in nucleotide substitutions in various States but there is no clear evidence of clustering of sequences according to their geog. origin. In practical terms, sequencing of viral RNA provides a means of testing the efficacy of further releases and subsequent spread of the virus if such a strategy is employed as a means of enhancing RHD as a biol. control of the wild rabbit in Australia.
RE.CNT 16 THERE ARE 16 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 54 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:112482 CAPLUS
DN 130:307345
TI Cloning and sequence analysis of two catechol-degrading gene clusters from the aniline-assimilating bacterium Frateuria species ANA-18
AU Murakami, Shuichiro; Takashima, Atsushi; Takemoto, Junji; Takenaka, Shinji; Shinke, Ryu; Aoki, Kenji
CS Laboratory of Applied Microbiology, Department of Biofunctional Chemistry. Faculty of Agriculture, Kobe University, Nada, Kobe, 657-8501, Japan
SO Gene (1999), 226(2), 189-198 CODEN: GENED6; ISSN: 0378-1119
PB Elsevier Science B.V.
DT Journal
LA English
AB The aniline-assimilating bacterium Frateuria species ANA-18 produced two catechol 1,2-dioxygenases, CD I and CD II, and two muconate cycloisomerases, MC I and MC II. The catA genes catA1 and catA2 encoding CD I and CD II, resp., were cloned from a gene library of this bacterium. The catA1 gene was clustered with catB1 encoding MC 1, catC1 encoding muconolactone isomerase (MI), catD encoding .beta.-ketoadipate enol-lactone hydrolase (ELH), and ORFR1 encoding a putative LysR-type regulator. The organization of these genes was ORFR1catB1C1D. The catA2 gene also constructed a gene cluster involving catB2 encoding MC II, catC2 encoding MI, and ORFR2 encoding a putative LysR-type regulator with the alignment of ORFR2catB2A2C2. The intergenic regions of ORFR1-catB1 and ORFR2-catB2 contained homologous sequences with the catR-catB intergenic region contg. a repression binding site and activation binding site of CatR in Pseudomonas putida. These findings suggest that the two cat clusters were regulated independently in their expression. When a product of cloned catD was added to a reaction mixt. contg. .beta.-ketoadipate enol-lactone, .beta.-ketoadipate was produced. This observation showed that the cloned catD encoded ELH and was expressed in Escherichia coli. We found that Frateuria sp. ANA-18 had a large plasmid with a mol. size more than 100 kb. Polymerase chain reaction amplifying partial catA genes and Southern hybridization analyses with probes contg. catA genes were conducted, to examine the localization of the two catA genes. We concluded that the catA1 and catA2 genes were located on the chromosomal and large plasmid DNAs, resp., in Frateuria sp. ANA-18.
RE.CNT 27 THERE ARE 27 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 55 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:64319 CAPLUS
DN 130:307296
TI Molecular epidemiology of Malaysian dengue 2 viruses isolated over twenty-five years (1968-1993)
AU Fong, M.-Y.; Koh, C.-L.; Lam, S.-K.
CS Department of Parasitology, University of Malaya, Kuala Lumpur, 50603, Malay.
SO Research in Virology (1998), 149(6), 457-464 CODEN: RESVEY; ISSN: 0923-2516
PB Editions Scientifiques et Medicales Elsevier
DT Journal
LA English
AB The limited sequencing approach was used to study the mol. epidemiol. of 24 Malaysian dengue 2 viruses which were isolated between 1968 and 1993. The sequences of a 240-nucleotide-long region across the envelope/non-structural 1 protein (E/NS1) gene junction of the isolates were detd. and analyzed. Alignment and comparison of the nucleotide and deduced amino acid sequences of the isolates revealed that nucleotide changes occurred mostly at the third position of a particular codon and were of the transition (A G, C U) type. Five nucleotide changes resulted in amino acid substitutions.

Pairwise comparisons of the nucleotide sequences gave divergence values ranging from 0 to 9.2%. At the amino acid level, the divergence ranged between 0 and 3.8%. Based on the 6% divergence as the cut-off point for genotypic classification , the isolates were grouped into two genotypes, I and II. Comparison of the nucleotide sequences of the Malaysian dengue isolates with those of the dengue viruses of other regions of the world revealed that members of genotypes I and II were closely related to viruses from the Indian Ocean and Western Pacific regions, resp.
RE.CNT 19 THERE ARE 19 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 56 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:50920 CAPLUS
DN 130:248605
TI Cloning of the gene for inorganic pyrophosphatase from a thermoacidophilic archaeon, Sulfolobus sp. strain 7, and overproduction of the enzyme by coexpression of tRNA for arginine rare codon
AU Wakagi, Takayoshi; Oshima, Tairo; Imamura, Hiromi; Matsuzawa, Hiroshi
CS Department of Biotechnology, The University of Tokyo, Tokyo, 113-8657, Japan
SO Bioscience, Biotechnology, and Biochemistry (1998), 62(12), 2408-2414 CODEN: BBBIEJ; ISSN: 0916-8451
PB Japan Society for Bioscience, Biotechnology, and Agrochemistry
DT Journal
LA English
AB The gene encoding an extremely stable inorg. pyrophosphatase from Sulfolobus sp. strain 7, a thermoacidophilic archaeon, was cloned and sequenced. An open reading frame consisted of 516 base pairs coding for a protein of 172-amino acid residues. The deduced sequence was supported by partial amino acid sequence analyses . All the catalytically important residues were conserved. A unique 17-base-pair sequence motif was found to be repeated four times in frame in the gene, encoding a cluster of acidic amino acids essential for the function. Although the codon usage of the gene was quite different from that of Escherichia coli, the gene was effectively expressed in E. coli. Coexpression of tRNAArg, cognate for the rare codon AGA in E. coli, however, further improved the prodn. of the enzyme, which occupied more than 85% of the sol. proteins obtained after removal of heat denatured E. coli proteins .
RE.CNT 32 THERE ARE 32 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 57 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1999:26518 CAPLUS
DN 130:207163
TI The hyperthermophilic bacterium Thermotoga maritima has two different classes of family C DNA polymerases: evolutionary implications
AU Huang, Yi-Ping; Ito, Junetsu
CS Department of Microbiology and Immunology, The University of Arizona, Tucson, AZ, 85724, USA
SO Nucleic Acids Research (1998), 26(23), 5300-5309 CODEN: NARHAD; ISSN: 0305-1048
PB Oxford University Press
DT Journal
LA English
AB Bacterial DNA polymerase III (family C DNA polymerase), the principal chromosomal replicative enzyme, is known to occur in at least three distinct forms which have provisionally been classified as class I (Escherichia coli DNA pol C-type), class II (Bacillus subtilis DNA pol C-type) and class III (cyanobacteria DNA pol C-type). We have identified two family C DNA polymerase sequences in the hyperthermophilic bacterium Thermotoga maritima. One DNA polymerase consisting of 842 amino acid residues and having a mol. wt. of 97 213 belongs to class I. The other one, consisting of 1367 amino acid residues and having a mol. wt. of 155 361, is a member of class II. Comparative sequence analyses suggest that the class II DNA polymerase is the principal DNA replicative enzyme of the microbe and that the class I DNA polymerase may be functionally inactive. A phylogenetic anal. using the class II enzyme indicates that T. maritima is closely related to the low G+C Gram-pos. bacteria, in particular to Clostridium acetobutylicum, and mycoplasmas. These results are in conflict with 16S rRNA-based phylogenies, which placed T. maritima as one of the deepest branches of the bacterial tree.
RE.CNT 74 THERE ARE 74 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 58 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:791955 CAPLUS
DN 130:205720
TI A histidine gene cluster of the hyperthermophile Thermotoga maritima: sequence analysis and evolutionary significance
AU Thoma, Ralf; Schwander, Martin; Liebl, Wolfgang; Kirschner, Kasper; Sterner, Reinhard
CS Abteilung fur Biophys. Chem., Biozentrum der Univ. Basel, Basel, CH-4056, Switz.
SO Extremophiles (1998), 2(4), 379-389 CODEN: EXTRFI; ISSN: 1431-0651
PB Springer-Verlag Tokyo
DT Journal
LA English
AB The sequences of histidine operon genes in hyperthermophiles are informative for understanding high protein thermostability and the evolution of metabolic pathways. Therefore, a cluster of eight his genes from the hyperthermophilic and phylogenetically early bacterium Thermotoga maritima was cloned and sequenced. The cluster has the gene order hisDCBdHAFI-E, lacking only hisG and hisBp, and does not contain intercistronic regions. This compact organization of his genes resembles the his operon of enterobacteria. Sequence anal. downstream of the stop codon of hisI-E identifies a region with a significantly higher cytosine over guanosine content, which is indicative of a rho-dependent termination of transcription of the his operon. Multiple sequence alignments of N1-((5'-phosphoribosyl)-formimino)-5-aminoimidazole-4-carboxyamide ribonucleotide isomerase (HisA) and of the cycloligase moiety of imidazoleglycerol phosphate synthase (HisF) support the previous assignment of (.beta..alpha.)5-barrel fold to these proteins . The alignments also reveal a second phosphate-binding motif located in the first halves of both enzymes and thereby support the hypothesis that HisA and HisF have evolved by a sequence of two gene duplication events. Comparison of the amino acid compns. of HisA and HisF from mesophiles and thermophiles shows that the thermostable variants of both enzymes contain a significantly increased no. of charged amino acid residues and may therefore be stabilized by addnl. salt bridges.

RE.CNT 54 THERE ARE 54 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 59 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:762758 CAPLUS
DN 130:120429
TI Can functional regions of proteins be predicted from their coding sequences? The case study of G- protein coupled receptors
AU Arrigo, P.; Fariselli, P.; Casadio, R.
CS Istituto Circuiti Elettronici, Consiglio Nazionale delle Richerche, Genoa, 1-161145, Italy
SO Gene (1998), 221(1), GC65-GC110 CODEN: GENED6; ISSN: 0378-1119
PB Elsevier Science B.V.
DT Journal
LA English
AB A filter based on a set of unsupervised neural networks trained with a winner-take-all strategy discloses signals along the coding sequences of G- protein coupled receptors. By comparing with the existing exptl. data it appears that these signals correlate with putative functional domains of the proteins . After protein alignment within subfamilies, signals cluster in protein regions which, according to the presently available exptl. results, are described as possible functional domains of the folded proteins . The mapping procedure reveals characteristic regions in the coding sequences common and/or characteristic of the receptor subtype. This is particularly noticeable for the third cytoplasmic loop, which is likely to be involved in the mol. coupling of all the subfamilies with G- proteins . The results indicate that our mapping can highlight intrinsic representative features of the coding sequences which, in the case of G- protein coupled receptors, are characteristic of protein functional regions and suggest a possible application of the filter for predicting functional determinants in proteins starting from the coding sequence.
RE.CNT 35 THERE ARE 35 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 60 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:622479 CAPLUS
DN 129:340945
TI Identification of major phylogenetic branches of inhibitory ligand-gated channel receptors
AU Xue, Hong
CS Department of Biochemistry, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong
SO Journal of Molecular Evolution (1998), 47(3), 323-333
CODEN: JMEVAU; ISSN: 0022-2844
PB Springer-Verlag New York Inc.
DT Journal
LA English
AB The gene superfamily of ligand-gated ion channel (LGIC) receptors is composed of members of excitatory LGIC receptors (ELGIC) and inhibitory LGIC receptors (ILGIC), all using amino acids as ligands. The ILGICs, including GABAA, Gly, and GluCl receptors, conduct Cl- when the ligand is bound. To evaluate the phylogenetic relationships among ILGIC members, 90 protein sequences were analyzed by both max.-parsimony and distance matrix-based methods. The strength of the resulting phylogenetic trees was evaluated by means of bootstrap. Four major phylogenetic branches are recognized. Branch I, called BZ, for the majority of the members are known to be related to benzodiazepine binding, is subdivided into IA, composed of all GABAA receptor .alpha. subunits, and IB, composed of the .gamma. and .epsilon. subunits, which are shown to be tightly linked. Branch II, named NB for non-benzodiazepine binding, and consisting of GABAA receptor .beta., .delta., .pi., and .rho. subunits, is further subdivided into IIA, contg. .beta. subunits; IIB, contg. .delta., and .pi. subunits; and IIC, contg. .rho. subunits. Branch IIIA, composed of vertebrate Gly receptors, is loosely clustered with Branch IIIB, composed of invertebrate GluCl receptors, to form Branch III, which is designated NA for being non-GABA responsive. Branch IV is called UD for being undefined in specificity. The existence of primitive forms of GABAA receptor non-.beta. subunits in invertebrates is first suggested by the present anal., and the identities of sequences p25123 from Drosophila melanogaster, s34469 from Lymnaea stagnalis, and u14635 and p41849 from Caenorhabditis elegans are detd. to be different from their previously given annotations. The proposed branching classification of ILGICs provides a phylogenetic map, based on protein sequences, for tracing the evolutionary pathways of ILGIC receptor subunits and detg. the identities of newly discovered subunits on the basis of their protein sequences.
RE.CNT 37 THERE ARE 37 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 61 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:541056 CAPLUS
DN 129:255795
TI The nos (nitrous oxide reductase) gene cluster from the soil bacterium Achromobacter cycloclastes: cloning, sequence analysis , and expression
AU McGuirl, Michele A.; Nelson, Laura K.; Bollinger, John A.; Chan, Yiu-Kwok; Dooley, David M.
CS Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT, 59717, USA
SO Journal of Inorganic Biochemistry (1998), 70(3,4), 155-169
CODEN: JIBIDJ; ISSN: 0162-0134
PB Elsevier Science Inc.
DT Journal
LA English
AB The nitrous oxide (N2O) reductase (nos) gene cluster from Archromobacter cycloclastes has been cloned and sequenced. Seven protein coding regions corresponding to nosR, nosZ (structural N2O reductase gene), nosD, nosF, nosY, nosL, and nosX are detected, indicating a genetic organization similar to that of Rhizobium meliloti. To aid homol. studies, nosR from R. meliloti has also been sequenced. Comparison of the deduced amino acid sequences with corresponding sequences from other organisms has also allowed structural and functional inferences to be made. The heterologous expression of NosD, NosZ (N2O reductase), and NosL is also reported. A model of the CuA site in N2O reductase, based on the crystal structure of this site in bovine heart cytochrome c oxidase, is presented. The model suggests that a His residue of the CuA domain may be a ligand to the catalytic CuZ site. In addn., the origin of the spectroscopically-obsd. Cys coordination to CuZ is discussed in terms of the sequence alignment of seven N2O reductases.
RE.CNT 53 THERE ARE 53 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 62 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:469798 CAPLUS
DN 129:227120

TI Self-organizing neural maps of the coding sequences of G-protein -coupled receptors reveal local domains associated with potentially functional determinants in the proteins
AU Arrigo, P.; Fariselli, P.; Casadio, R.
CS Istituto Circuiti Elettronici, Consiglio Nazionale delle Ricerche, Genoa, I-16149, Italy
SO Proceedings International Conference on Intelligent Systems for Molecular Biology, 5th, Halkidiki, Greece, June 21-25, 1997 (1997), 44-47. Editor(s): Gaasterland, Terry. Publisher: AAAI Press, Menlo Park, Calif. CODEN: 66LJAU
DT Conference
LA English
AB Mapping of the coding sequences of the best characterized subfamilies of G- protein -coupled receptors is performed with unsupervised neural networks based on a winner-take-all strategy. High order features therefrom extd. originate signals along the aligned protein sequences of the different subfamilies. These plots reveal characteristic domains common and/or characteristic of the receptor subfamily. By comparison with the existing exptl. results, it is obtained that most of the regions signalled by clustering overlap with possible functional regions in the folded proteins . This is particularly noticeable for the third cytoplasmic loop, which is likely to be involved in mol. coupling with the G- proteins . The results suggest that functional regions in proteins may be characterized by intrinsic representative features in the coding sequences which can be enlightened by high order mapping.
RE.CNT 19 THERE ARE 19 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 63 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:395923 CAPLUS
DN 129:158083
TI Prediction of functional residues in water channels and related proteins
AU Froger, A.; Tallur, B.; Thomas, D.; Delamarche, C.
CS UPRES-A CNRS 6026, Biologie Cellulaire et Reproduction, Equipe "Canaux et Recepteurs Membranaires," Universite de Rennes 1, Rennes, 35042, Fr.
SO Protein Science (1998), 7(6), 1458-1468 CODEN: PRCIEI; ISSN: 0961-8368
PB Cambridge University Press
DT Journal
LA English
AB In this paper, we present an updated classification of the ubiquitous MIP (Major Intrinsic Protein) family proteins , including 153 fully or partially sequenced members available in public databases. Presently, about 30 of these proteins have been functionally characterized, exhibiting essentially two distinct types of channel properties: (1) specific water transport by the aquaporins, and (2) small neutral solutes transport, such as glycerol by the glycerol facilitators. Sequence alignments were used to predict amino acids and motifs discriminant in channel specificity. The protein sequences were also analyzed using statistical tools (comparisons of means and correspondence anal.). Five key positions were clearly identified where the residues are specific for each functional subgroup and exhibit high dissimilar physico-chem. properties. Moreover, we have found that the putative channels for small neutral solutes clearly differ from the aquaporins by the amino acid content and the length of predicted loop regions, suggesting a substrate filter function for these loops. From these results, we propose a signature pattern for water transport.
RE.CNT 54 THERE ARE 54 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 64 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:373877 CAPLUS
DN 129:158055
TI Computer analysis of amino acid sequences: the case of plant virus capsid proteins
AU Koonin, Eugene V.; Mushegian, Arcady R.; Dolja, Valerian V.
CS National Center for Biotechnology Information, Computational Biology Branch, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA
SO Methods in Molecular Biology (Totowa, New Jersey) (1998), 81(Plant Virology Protocols), 319-337 CODEN: MMBIED; ISSN: 1064-3745
PB Humana Press Inc.
DT Journal
LA English
AB The field of computer anal. of protein sequences has already become very diverse. Here we present only a small set of relatively straightforward, statistically reliable techniques that allow a researcher to rapidly progress from an uncharacterized protein sequence to a meaningful multiple alignment and or conserved motifs useful both for the purpose of classification and exptl. design. What has to be emphasized is the interaction between basic methods for database screening in search of pairwise sequence similarity (e.g. BLAST), multiple alignment construction methods (e.g. MACAW) and motif anal. methods (e.g. MoST) and iterative application of each of these approaches. All the programs discussed run under the UNIX operation system, typically on both Suns and SGI computers, except for MACAW which runs on PC and MAC platforms.
RE.CNT 50 THERE ARE 50 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 65 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:373426 CAPLUS
DN 129:118524
TI Genomic structure and sequence analysis of human HOXA-9
AU Kim, Myoung Hee; Chang, Hwa-Hyoung; Shin, Chuog; Cho, Myungsun; Park, Dalkeun; Park, Hyoung Woo
CS Department of Anatomy, Yonsei Univ. College of Medicine, Seoul, S. Korea
SO DNA and Cell Biology (1998), 17(5), 407-414 CODEN: DCEBE8; ISSN: 1044-5498
PB Mary Ann Liebert, Inc.
DT Journal
LA English
AB In order to understand the regulatory mechanisms establishing and maintaining HOXA-9 gene expression, structural information about the gene is a prerequisite. Therefore, we sequenced the 7.2-kb region of the human HOXA-9 gene and mapped the positions of two partial cDNAs consisting of one of two 5' exons, AB (358 bp) or CD (568 bp), and a common 3' exon (exon II), which are sepd. by 5.4- and 1.0-kb introns, resp. When the amino acid sequence homologies were compared with those of other Hox genes belonging to the same paralogous group, exon CD exhibited the strongest homol.: 73% of 91 aa residues exactly matched those of chicken HOXA-9. An intermediate exon (90 bp) was detected within exon CD. It was surrounded by a splice acceptor and a donor at both the 5' and 3' ends, and one

branchpoint site was found near the splice acceptor site. Nucleotide sequence anal. along this region revealed two TATA boxes, one CAAT box, one GC box, and one each of the following binding sites-engrailed, eve-stripe2-hb3, and Krox20- just upstream of exon CD. A CpG island and two RARE repeats were detected within intron I. Northern blot anal. showed that at least four main transcripts were generated along this region: all fetal tissues tested (brain, lung, liver, and kidney) produced a 1.8-kb homeobox-contg. transcript (HA-9A); a 2.2- and a 3.3- kb transcript were generated from exon CD and exon II (HA-9B), esp. in fetal and adult kidneys as well as in adult skeletal muscle; the 1.0-kb transcript was likely to be generated by the intermediate exon in all adult and fetal tissues. Several weak bands without tissue specificity were likely to be contributed by the hybrid transcripts between HOXA-9 and the other HOXA gene(s). Together, these results may account for the unique degree of conservation of the HOX cluster in general.
RE.CNT 22 THERE ARE 22 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 66 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:339405 CAPLUS
DN 129:91234
TI Cloning, sequence analysis and expression of the basidiomycete Lentinus edodes gene uck1, encoding UMP-CMP kinase, the homolog of Saccharomyces cerevisiae URA6 gene
AU Kaneko, Shinya; Miyazaki, Yasumasa; Yasuda, Toru; Shishido, Kazuo
CS Dep. of Life Science, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama, 226-8501, Japan
SO Gene (1998), 211(2), 259-266 CODEN: GENED6; ISSN: 0378-1119
PB Elsevier Science B.V.
DT Journal
LA English
AB Sequence anal. of the downstream region of the basidiomycete Lentinus edodes priB gene encoding a protein with a 'Zn(II)2Cys6 zinc cluster ' DNA-binding motif suggested the presence of a Saccharomyces cerevisiae URA6 gene homolog encoding UMP kinase. We isolated a corresponding cDNA from a mature fruiting-body cDNA library of L. edodes. The nucleotide sequence of this was detd. and compared with that of the genomic DNA, revealing that the URA6 gene homolog encodes 277 amino acids (aa) and is interrupted by four small introns. The deduced aa sequence showed an overall identity of 51.1% to that of the S. cerevisiae URA6 gene product. The URA6 homolog protein produced in Escherichia coli using the glutathione S-transferase gene fusion system was found to catalyze the phosphoryl transfer from ATP to UMP and CMP efficiently and also to AMP and dCMP with lower efficiencies. Thus, the URA6 gene homolog was designated uck1 and its product UMP-CMP kinase. Northern-blot anal. showed that the uck1 is actively transcribed in the gill tissue of mature fruiting bodies of L. edodes, implying that uck1 may play a role during the formation of basidiospores occurs in the gill tissue.
RE.CNT 37 THERE ARE 37 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 67 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:243446 CAPLUS
DN 128:290995
TI Structural and functional implications of sequence diversity of Pseudomonas aeruginosa genes oriC, ampC, and fliC
AU Spangenberg, Claudia; Montie, Thomas C.; Tummler, Burkhard
CS Klinische Forschergruppe, Zentrum Biochemie, Medizinische Hochschule Hannover, Hannover, D-30623, Germany
SO Electrophoresis (1998), 19(4), 545-550 CODEN: ELCTDN; ISSN: 0173-0835
PB Wiley-VCH Verlag GmbH
DT Journal
LA English
AB Sequence anal. of 3 representative gene loci, oriC, ampC, and fliC, in 19 P. aeruginosa strains revealed a low sequence diversity that does not correlate with the extensive diversity of P. aeruginosa habitats. Single point mutations lead to a sequence diversity of 0.40%, 0.38%, and 0.59% for oriC, ampC, and a-type fliC, resp., but of only 0.05% for b-type flagellin genes. The analyzed genes encode highly conserved functions that are subject to strong selective pressure. The detected nucleotide substitutions of oriC, accumulating in a central 95-bp region, affect neither the putative DnaA binding sites nor the 13-bp direct repeats that presumably provide the sites to open oriC duplex DNA. Even in P. aeruginosa strain DSM 1128, which exhibits an unusually high sequence variability in several analyzed genes, the 9-bp and 13-bp motifs are conserved, reflecting their essential functional role in replication initiation. The 2 flagellin types, differing by 37-38% in their primary structure, exhibit pronounced structural and functional homol., as shown by alignment of flagellin variants by hydrophobicity index, probability of surface exposure, chain flexibility and antigenicity, and by cross-reactivity between both proteins using specific antisera. 5 Nonsynonymous nucleotide substitutions of ampC lead to .beta.-lactamase variants that differ in recognition and turnover of substrate, as deduced from the 3-dimensional structure of the highly homologous Enterobacter cloacae .beta.-lactamase and confirmed by inhibition kinetics. The identified point mutations in the 3 genes are classified as selectively equiv. sequence variants indicating neutral genetic drift as a mechanism of mol. evolution in P. aeruginosa, rather than pos. selection.

L9 ANSWER 68 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:238074 CAPLUS
DN 129:64721
TI Source and target enzyme signature in serine protease inhibitor active site sequences
AU Prakash, Balaji; Murthy, M. R. N.
CS Molecular Biophysics Unit, Indian Institute of Science, Bangalore, 560 012, India
SO Journal of Biosciences (Bangalore, India) (1997), 22(5), 555-565 CODEN: JOBSDN; ISSN: 0250-5991
PB Indian Academy of Sciences
DT Journal
LA English
AB Amino acid sequences of proteinaceous proteinase inhibitors have been extensively analyzed for deriving information regarding the mol. evolution and functional relationship of these proteins . These sequences have been grouped into several well-defined families. It was found that the phylogeny constructed with the sequences corresponding to the exposed loop responsible for inhibition has several branches that resemble those obtained from comparisons using the entire sequence. The major branches of the unrooted tree corresponded to the families to which the inhibitors belonged. Further branching is related to the

enzyme specificity of the inhibitor. Examn. of the active site loop sequences of trypsin inhibitors revealed that there are strong preferences for specific amino acids at different positions of the loop. These preferences are inhibitor class specific. Inhibitors active against more than one enzyme occur within a class and confirm to class-specific sequence in their loops. Hence, only a few positions in the loop seem to det. the specificity. The ability to inhibit the same enzyme by inhibitors that belong to different classes appears to be a result of convergent evolution.
RE.CNT 7 THERE ARE 7 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 69 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:207712 CAPLUS
DN 129:13761
TI Sequence analysis of glutamate dehydrogenase (GDH) from the hyperthermophilic archaeon Pyrococcus sp. KOD1 and comparison of the enzymic characteristics of native and recombinant GDHs
AU Rahman, R. N. Z. A.; Fujiwara, S.; Takagi, M.; Imanaka, T.
CS Dep. Synthetic Chemistry and Biol. Chem., Graduate Sch. Eng., Kyoto Univ., Kyoto, 606-01, Japan
SO Molecular & General Genetics (1998), 257(3), 338-347
CODEN: MGGEAE; ISSN: 0026-8925
PB Springer-Verlag
DT Journal
LA English
AB The gdhA gene encoding glutamate dehydrogenase (GDH) from the hyperthermophilic archaeon Pyrococcus sp. KOD1 was cloned and sequenced. Phylogenetic anal. was performed on an alignment of 25 GDH sequences including KOD1-GDH, and two protein families were distinguished, as previously reported. KOD1-GDH was classified as new member of the hexameric GDH Family II. The gdhA gene was expressed in Escherichia coli, and recombinant KOD1-GDH was purified. Its enzymic characteristics were compared with those of the native KOD1-GDH. Both enzymes had a mol. mass of 47,300 Da and were shown to be functional in a hexameric form (284 kDa). The N-terminal amino acid sequences of native KOD1-GDH and the recombinant GDH were VEIDPFEMAV and MVEIDPFEMA, resp., indicating that native KOD1-GDH does not retain the initial methionine at the N-terminus. The recombinant GDH displayed enzyme characteristics similar to those of the native GDH, except for a lower level of thermostability, with a half-life of 2 h at 100.degree., compared to 4 h for the native enzyme purified from KOD1. Kinetic studies suggested that the reaction is biased towards glutamate prodn. KOD1-GDH utilized both coenzymes NADH and NADPH, as do most eukaryal GDHs.
RE.CNT 58 THERE ARE 58 CITED REFERENCES AVAILABLE FOR THIS RECORD ALL CITATIONS AVAILABLE IN THE RE FORMAT

L9 ANSWER 70 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:153145 CAPLUS
DN 128:290651
TI Rolling-circle plasmids from Bacillus subtilis: complete nucleotide sequences and analyses of genes of pTA1015, pTA1040, pTA1050 and pTA1060, and comparisons with related plasmids from Gram-positive bacteria
AU Meijer, Wilfried J. J.; Wisman, G. Bea A.; Terpstra, Peter; Thorsted, Peter B.; Thomas, Chris M.; Holsappel, S.; Venema, Gerard; Bron, Sierd
CS Kerklaan 30, Department of Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, Haren, 9751 NN, Neth.
SO FEMS Microbiology Reviews (1998), 21(4), 337-368
CODEN: FMREE4; ISSN: 0168-6445
PB Elsevier Science B.V.
DT Journal; General Review
LA English
AB A review, with 96 refs. Most small plasmids of Gram-pos. bacteria use the rolling-circle mechanism of replication and several of these have been studied in considerable detail at the DNA level and for the function of their genes. Although most of the common lab. Bacillus subtilis 168 strains do not contain plasmids, several industrial strains and natural soil isolates do contain rolling circle replicating (RCR) plasmids. So far, knowledge about these plasmids was mainly limited to: (i) a classification into seven groups, based on size and restriction patterns; and (ii) DNA sequences of the replication region of a limited no. of them. To increase the knowledge, also with respect to other functions specified by these plasmids, we have detd. the complete DNA sequence of four plasmids, representing different groups, and performed computer-assisted and exptl. analyses on the possible function of their genes. The plasmids analyzed are pTA1015 (5.8 kbp), pTA1040 (7.8 kbp), pTA1050 (8.4 kbp), and pTA1060 (8.7 kbp). These plasmids have a structural organization similar to most other known RCR plasmids. They contain highly related replication functions, both for leading and lagging strand synthesis. Plasmids pTA1015 and pTA1060 contain a mobilization gene enabling their conjugative transfer. Strikingly, in addn. to the conserved replication modules, these plasmids contain unique module(s) with genes which are not present on known RCR plasmids of other Gram-pos. bacteria. Examples are genes encoding a type I signal peptidase and genes encoding proteins belonging to the family of response regulator aspartate phosphatases. The latter are likely to be involved in the regulation of post-exponential phase processes. The presence of these modules on plasmids may reflect an adaptation to the special conditions to which the host cells were exposed.

L9 ANSWER 71 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:88264 CAPLUS
DN 128:226856
TI Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6
AU Ansari-Lari, M. Ali; Oeltjen, John C.; Schwartz, Scott; Zhang, Zheng; Muzny, Donna M.; Lu, Jing; Gorrell, James H.; Chinault, A. Craig; Belmont, John W.; Miller, Webb; Gibbs, Richard A.
CS Department Molecular and Human Genetics, Baylor College Medicine, Houston, TX, 77030, USA
SO Genome Research (1998), 8(1), 29-40 CODEN: GEREFS; ISSN: 1088-9051
PB Cold Spring Harbor Laboratory Press
DT Journal
LA English
AB The Human Genome Project has created a formidable challenge: the extn. of biol. information from extensive amts. of raw sequence. With the increasing availability of genomic sequence from other species, one approach to extg. coding and regulatory element information is through cross-species sequence comparison. To assess the strengths and weaknesses of this methodol. for large-scale sequence anal., 227 kb of mouse sequence syntenic to a gene-rich cluster on human chromosome 12p13 was obtained. Primarily through

percent identity plots (PIPs) of SIM comparative sequence alignments , the sequence of coding reigns, putative alternative exons, conserved noncoding regions, and correlation in repetitive element insertions were easily detd. The anal. demonstrated that the no., order, and orientation of all 17 genes are conserved between the two species, whereas two human pseudogenes are absent in mouse. In addn., apart from MIRs, no direct correlation of distribution or position of the majority of repetitive elements between the two species is seen. Finally, in examg. the synonymous and nonsynonymous substitution rates in the conserved genes, a large variation in nonsynonymous rats is obsd. indicating that the genes in this region are diverging at different rates. This study indicates the utility and strength of large-scale cross-species sequence comparisons in the extn. of biol. information from raw sequence, esp. when combined with other computational tools such as GRAIL and BLAST.

L9 ANSWER 72 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1998:80572 CAPLUS
DN 128:227894
TI A structural model of picornavirus leader proteinases based on papain and bleomycin hydrolase
AU Skern, Tim; Fita, Ignasi; Guarne, Alba
CS Medical Faculty, Institute Biochemistry, University Vienna, Vienna, A-1030, Austria
SO Journal of General Virology (1998), 79(2), 301-307
CODEN: JGVIAY; ISSN: 0022-1317
PB Society for General Microbiology
DT Journal
LA English
AB The leader (L) proteinases of aphthoviruses (foot-and-mouth disease viruses) and equine rhinovirus serotypes 1 and 2 cleave themselves from the growing polyprotein. This cleavage occurs intramolecularly between the C terminus of the L proteinases and the N terminus of the subsequent protein VP4. The foot-and-mouth disease virus enzyme has been shown, in addn., to cleave at least one cellular protein , the eukaryotic initiation factor 4G. Mechanistically, inhibitor studies and sequence anal. have been used to classify the L proteinases as papain-like cysteine proteinases . However, sequence identity within the L proteinases themselves is low (between 18% and 32%) and only 14% between the L proteinases and papain. Secondary structure predictions, sequence alignments that take into account the positions of the essential catalytic residues, and structural considerations have been used in this study to investigate more closely the relationships between the L proteinases and papain. In spite of the low sequence identities, the analyses strongly suggest that the L proteinases of foot-and-mouth disease virus and of equine rhinovirus 1 have similar overall fold to that of papain. Regions in the L proteinases corresponding to all five .alpha.-helixes and seven .beta.-sheets of papain could be identified. Further comparisons with the proteinase bleomycin hydrolase, which also displays a papain topol. in spite of important differences in size and amino acid sequence, support these conclusions and suggest how a C-terminal extension, present in all three L proteinases , and predicted to be an .alpha.-helix, might enable C-terminal self-processing to occur.

L9 ANSWER 73 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1997:795541 CAPLUS
DN 128:111413
TI Sequence and phylogenetic analyses of HIV-1 infection in Vietnam: subtype E in commercial sex workers (CSW) and injection drug users (IDU)
AU Nerurkar, Vivek R.; Nguyen, Hien Tran; Woodward, Cora L.; Hoffmann, Peter R.; Dashwood, Wan-Mohaiza; Long, Hoang Thuy; Morens, David M.; Detels, Roger; Yanagihara, Richard
CS Retrovirology Research Laboratory, Pacific Biomedical Research Center, Leahi Hospital, University of Hawaii at Manoa, Honolulu, HI, 96816, USA
SO Cellular and Molecular Biology (Paris) (1997), 43(7), 959-968 CODEN: CMOBEF; ISSN: 0145-5680
PB C.M.B. Association
DT Journal
LA English
AB More than 4,000 persons with human immunodeficiency virus type 1 (HIV-1) infection have been identified in Vietnam through sentinel surveillance since 1990, when the first case of HIV-1 infection was diagnosed in a young woman in Ho Chi Minh City. Currently, the estd. HIV-1 seroprevalences of 10% for injection drug users (IDU) and 3% for female com. sex workers (CSW) in Vietnam are comparable to those obsd. in the same risk groups in Thailand five years ago. To clarify if concurrent epidemics with different HIV-1 subtypes (or clades) are occurring among different high-risk behavior groups in Vietnam, we conducted a genotypic anal. of HIV-1 by amplifying and sequencing a 325-nucleotide region spanning the principal neutralizing domain, or V3 loop, of the gp120-encoding env gene from genomic DNA extd. from dried, filter paper-blotted blood samples, collected in Apr./May and August/Sept. 1995 from 8 HIV-1-seropos. CSW in Ho Chi Minh City, Can Tho and An Giang provinces and from 16 IDU in Ho Chi Minh City, Hanoi, Nha Trang and An Giang province. Sequence alignment and comparison with other HIV-1 subtypes indicated that the HIV-1 strains from CSW and IDU in Vietnam were genetically most similar to subtype E strains from Cambodia. The interstrain genetic variation among the Vietnam HIV-1 env sequences ranged from 0.3% to 9.0% (mean, 4.6%). Phylogenetic anal. verified that some of the Vietnam HIV-1 strains formed discrete clusters and were indistinguishable from other Southeast Asian strains. The demonstration of subtype E in both CSW and IDU in Vietnam contrasts sharply with the previously obsd. HIV-1 clade restriction in these high-risk behavior groups in nearby Thailand.

L9 ANSWER 74 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1997:607040 CAPLUS
DN 127:262278
TI Visual BLAST and visual FASTA: graphic workbenches for interactive analysis of full BLAST and FASTA outputs under Microsoft Windows 95/NT
AU Durand, P.; Canard, L.; Mornon, J. P.
CS Systems Mol. Biol. Structurale, Lab. Mineralogie-Cristallographie, Univ. Paris VI-Paris VII, Paris, F-75252, Fr.
SO CABIOS, Computer Applications in the Biosciences (1997), 13(4), 407-413 CODEN: COABER; ISSN: 0266-7061
PB Oxford University Press
DT Journal
LA English
AB When routinely analyzing protein sequences, detailed anal. of database search results made with BLAST and FASTA becomes exceedingly time consuming and tedious work, as the resultant file may contain a list of hundreds of potential homologies. The interpretation of these results is usually carried out with a text editor which is not a convenient tool for this anal. In addn., the format of data within BLAST and FASTA output files makes them difficult to read. To facilitate and accelerate this anal., we present, for the first time, two easy-to-use programs designed for interactive anal. of full

BLAST and FASTA output files contg. protein sequence alignments . The programs, Visual BLAST and Visual FASTA, run under Microsoft Windows 95 or NT systems. They are based on the same intuitive graphical user interface (GUI) with extensive viewing, searching, editing, printing and multithreading capabilities. These programs improve the browsing of BLAST/FASTA results by offering a more convenient presentation of these results. They also implement on a computer several anal. tools which automate a manual methodol. used for detailed anal. of BLAST and FASTA outputs. These tools include a pairwise sequence alignment viewer, a Hydrophobic Cluster anal. plot alignment viewer and a tool displaying a graphical map of all database sequences aligned with the query sequence. In addn., Visual Blast includes tools for multiple sequence alignment anal. (with an amino acid patterns search engine), and Visual FASTA provides a GUI to the FASTA program.

L9 ANSWER 75 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1997:143869 CAPLUS
DN 126:207884
TI Progressive multiple alignment with constraints
AU Myers, Gene; Selznick, Sanford; Zhang, Zheng; Miller, Webb
CS Department of Computer Science, University of Arizona, Tucson, AZ, 85721, USA
SO Journal of Computational Biology (1996), 3(4), 563-572
CODEN: JCOBEM; ISSN: 1066-5277
PB Liebert
DT Journal
LA English
AB A progressive alignment algorithm produces a multialignment of a set of sequences by repeatedly aligning pairs of sequences and/or previously generated alignments . We describe a method for guaranteeing that the alignment generated by a progressive alignment strategy satisfies a user-specified collection of constraints about where certain sequence positions should appear relative to others. Our main result is an algorithm to compute just the "prime" constraints that are implied by the user-given constraints; these are shown to be precisely the constraints that the alignment algorithm must obey. In practice, the time required to handle constraints is negligible and frequently much less than the time saved because the constraints permit searching a restricted region of the dynamic-programming grid. An alignment of the .beta.-like globin gene cluster of several mammals illustrates the practically of the method.

L9 ANSWER 76 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1997:106799 CAPLUS
DN 126:167184
TI Cloning and sequence analysis of genes encoding xylanases and acetyl xylan esterase from Streptomyces thermoviolaceus OPC-520
AU Tsujibo, Hiroshi; Ohtsuki, Takahito; IiO, Takashi; Yamazaki, Isao; Miyamoto, Katsushiro; Sugiyama, Masanori; Inamori, Yoshihiko
CS Osaka University of Pharmaceutical Sciences, Osaka, 569-11, Japan
SO Applied and Environmental Microbiology (1997), 63(2), 661-664 CODEN: AEMIDF; ISSN: 0099-2240
PB American Society for Microbiology
DT Journal
LA English
AB Three genes encoding two types of xylanases (STX-I and STX-II) and an acetyl xylan esterase (STX-III) from Streptomyces thermoviolaceus OPC-520 were cloned, and their DNA sequences were detd. The nucleotide sequences showed that genes stx-II and stx-III were clustered on the genome. The stx-I, stx-II, and stx-III genes encoded deduced proteins of 51, 35.2, and 34.3 kDa, resp. STX-I and STX-II bound to both insol. xylan and cryst. cellulose (Avicel). Alignment of the deduced amino acid sequences encoded by stx-I, stx-II, and stx-III demonstrated that the three enzymes contain two functional domains, a catalytic domain and a substrate-binding domain. The catalytic domains of STX-I and STX-II showed high sequence homol. to several xylanases which belong to families F and G, resp., and that of STX-III showed striking homol. with an acetyl xylan esterase from S. lividans, nodulation proteins of Rhizobium sp., and chitin deacetylase of Mucor rouxii. In the C-terminal region of STX-I, there were three reiterated amino acid sequences starting from C-L-D, and the repeats were homologous to those found in xylanase A from S. lividans, coagulation factor G subunit .alpha. from the horseshoe crab, Rarobacter faecitabidus protease I, .beta.-1,3-glucanase from Oerskovia xanthineolytica, and the ricin B chain. However, the repeats did not show sequence similarity to any of the nine known families of cellulose-binding domains (CBDs). On the other hand, STX-II and STX-III contained identical family II CBDs in their C-terminal regions.

L9 ANSWER 77 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1997:101186 CAPLUS
DN 126:326331
TI Cloning, sequencing, and developmental expression of phosphofructokinase from Dictyostelium discoideum
AU Estevez, Antonio M.; Martinez-Costa, Oscar H.; Sanchez, Valentina; Aragon, Juan J.
CS Facultad Medicina, Universidad Autonoma Madrid, Madrid, E-28029, Spain
SO European Journal of Biochemistry (1997), 243(1/2), 442-451 CODEN: EJBCAI; ISSN: 0014-2956
PB Springer
DT Journal
LA English
AB Phosphofructokinase (PFK) from D. discoideum is a non-allosteric enzyme that lacks any of the characteristic regulatory mechanisms of PFK from other cells. We have detd. the DNA sequence and analyzed the amino acid sequence of D. discoideum PFK, as an initial step toward understanding the peculiar properties of this enzyme. Three overlapping fragments, 2 of cDNA and 1 of genomic DNA, were isolated, which together could encode the complete sequence of D. discoideum PFK. The constructed full-length cDNA coded for a protein of 834 amino acids, with a calcd. mol. mass of 92.4 kDa, which was similar to other eukaryotic and prokaryotic PFK. Alignments of the amino acid sequence with other isoenzymes revealed that many of the amino acid residues assigned to binding sites of substrates and allosteric effectors are conserved in this enzyme, but changes were also found that may contribute to the absence of allosteric mechanisms. A phylogenetic tree for the eukaryotic PFK family was constructed and showed that the N-terminal domain clustered with those of yeast subunits, whereas the C-terminal domain was more related to PFK from metazoa. Southern blotting indicated that D. discoideum PFK is encoded by a single gene. The enzyme is present throughout the life cycle of D. discoideum, with a gradual decrease of its expression during development.

L9 ANSWER 78 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1997:85366 CAPLUS
DN 126:209417

TI Antigenic and molecular analyses of the variability of bovine respiratory syncytial virus G glycoprotein
AU Prozzi, Deborah; Walravens, Karl; Langedijk, Johannes P. M.; Daus, Franz; Kramps, Johannes A.; Letesson, Jean-Jacques
CS Unite de Microbiologie et Immunologie, Facultes Univ. Notre Dame de la Paix, Namur, 5000, Belg.
SO Journal of General Virology (1997), 78(2), 359-366
CODEN: JGVIAY; ISSN: 0022-1317
PB Society for General Microbiology
DT Journal
LA English
AB Antigenic variation among eight bovine respiratory syncytial virus (BRSV) isolates was detd. using monoclonal antibodies (MAbs) specific for the attachment (G) protein . Two major (and one intermediate) subgroups were identified, as well as one strain that did not fit any pattern. The subgroups could also be differentiated on the basis of the Mr of the F protein cleavage product, F2. The nucleotide sequence of the G gene of seven of the BRSV strains was detd. and compared with published G gene sequences. Subgroups A and A/B were more closely related in protein sequence than subgroups A and B or subgroups A/C and B. These results could not be correlated with those obtained by the detn. of the Mr of the F2 polypeptide . Multiple sequence alignments showed a high level of aminio acid identity at the inter-subgroup level (85% identity between subgroup A and subgroup B strains), similar to the intra-subgroup human (H)RSV identity, suggesting that the BRSV isolates form a continuum rather than distinct subgroups. However, unusual variability was obsd. within the immunodominant domain (amino acids 174-188) in contrast with the situation in HRSV strains belonging to the same subgroup.

L9 ANSWER 79 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:684585 CAPLUS
DN 125:321061
TI A computer aided system for systematic production and revision of sequence patterns
AU Ripoche, H.; Sallantin, J.
CS LIRMM, Montpellier, 34392, Fr.
SO Biochimie (1996), 78(5), 370-375 CODEN: BICMBE; ISSN: 0300-9084
PB Elsevier
DT Journal
LA English
AB Two complementary fields, object-oriented databases and machine learning, were used to produce and revise a set of protein sequence patterns. First, object-oriented query languages were shown to be well suited for the prodn. of patterns as well as for the interpretation of the biol. function of new (uncharacterized) sequences. Next, a classification was built from the set of sequences according to the pattern matches. This classification may be criticized by a specific anal. method, which yields back to revise sequences and patterns. In this application, concept lattices were used as a classification method and sequence multiple alignment for criticism.

L9 ANSWER 80 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:617709 CAPLUS
DN 125:268890
TI A new Azotobacter vinelandii mannuronan C-5-epimerase gene (algG) is part of an alg gene cluster physically organized in a manner similar to that in Pseudomonas aeruginosa
AU Rehm, Bernd H.; Ertesvag, Helga; Valla, Svein
CS Lehrstuhl Mikrobiologie Mikroorganismen, Ruhr-Univ. Bochum, Bochum, 44780, Germany
SO Journal of Bacteriology (1996), 178(20), 5884-5889
CODEN: JOBAAY; ISSN: 0021-9193
PB American Society for Microbiology
DT Journal
LA English
AB Alginate is an unbranched polysaccharide composed of the two sugar residues .beta.-D-mannuronic acid (M) and .alpha.-L-guluronic acid (G). The M/G ratio and sequence distribution in alginates vary and are of both biol. and com. significance. The authors have previously shown that a family of highly related mannuronan C-5-epimerase genes (algE1 to -E5) controls these parameters in Azotobacter vinelandii, by catalyzing the Ca2+-dependent conversion of M to G at the polymer level. In this report, the authors describe the cloning and expression of a new A. vinelandii epimerase gene (here designated algG), localized 29 nucleotides downstream of the previously described gene algJ. Sequence alignments show that algG does not belong to the same class of genes as algE1 to -E5 but that it share 66% sequence identity with a previously described mannuronan C-5-epimerase gene (also designated algG) from Pseudomonas aeruginosa. A. vinelandii algG was expressed in Escherichia coli, and the enzyme was found to catalyze epimerization in the absence of Ca2+, although the presence of the cation stimulated the activity moderately. Surprisingly, all activity was blocked by Zn2+. P. aeruginosa AlgG has been reported to contain an N-terminal export signal sequence which is cleaved off during expression in E. coli. This does not happen with A. vinelandii AlgG, which appears to be produced at least partly in an insol. form when expressed at high levels in E. coli. DNA sequencing analyses of the regions flanking algG suggest that the gene is localized in a cluster of genes putatively involved in alginate biosynthesis, and the organization of this cluster appears to be the same as previously described for P. aeruginosa.

L9 ANSWER 81 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:590680 CAPLUS
DN 125:268847
TI Comparison of a .beta.-glucosidase and a .beta.-mannosidase from the hyperthermophilic archaeon Pyrococcus furiosus. Purification, characterization, gene cloning, and sequence analysis
AU Bauer, Michael W.; Bylina, Edward J.; Swanson, Ronald V.; Kelly, Robert M.
CS Dep. Chem. Eng., North Carolina State Univ., Raleigh, NC, 27695-7905, USA
SO Journal of Biological Chemistry (1996), 271(39), 23749-23755 CODEN: JBCHA3; ISSN: 0021-9258
PB American Society for Biochemistry and Molecular Biology
DT Journal
LA English
AB Two distinct exo-acting, .beta.-specific glycosyl hydrolases were purified to homogeneity from crude cell exts. of the hyperthermophilic archaeon Pyrococcus furiosus: a .beta.-glucosidase, corresponding to the one previously purified by Kengen et al. (Kengen, S. W. M., Luesink, E. J., Stams, A. J. M., and Zehnder, A. J. B. (1993) Eur. J. Biochem. 213, 305-312), and a .beta.-mannosidase. The .beta.-mannosidase and .beta.-glucosidase genes were isolated from a genomic library by expression screening. The nucleotide sequences predicted polypeptides with 510 and 472 amino acids corresponding to calcd. mol. masses of 59.0 and 54.6 kDa for the .beta.-mannosidase and the .beta.-glucosidase, resp. The .beta.-glucosidase gene was identical to that reported by Voorhorst et al. (Voorhorst, W. G. B., Eggen, R. I. L., Luesink, E. J., and

deVos, W. M. (1995) J. Bacteriol. 177, 7105-7111; GenBank accession no. U37557). The deduced amino acid sequences showed homol. both with each other (46.5% identical) and with several other glycosyl hydrolases, including the .beta.-glycosidases from Sulfolobus solfataricus, Thermotoga maritima, and Caldocellum saccharolyticum. Based on these sequence similarities, the .beta.-mannosidase and the .beta.-glucosidase can both be classified as family 1 glycosyl hydrolases. In addn., the .beta.-mannosidase and .beta.-glucosidase from P. furiosus both contained the conserved active site residues found in all family 1 enzymes. The .beta.-mannosidase showed optimal activity at pH 7.4 and 105.degree.C. Although the enzyme had a half-life of greater than 60 h at 90.degree.C, it is much less thermostable than the .beta.-glucosidase, which had a reported half-life of 85 h at 100.degree.C. Km and Vmax values for the .beta.-mannosidase were detd. to be 0.79 mM and 31.1 .mu.mol para-nitrophenol released/min/mg with p-nitrophenyl-.beta.-D-mannopyranoside as substrate. The catalytic efficiency of the .beta.-mannosidase was significantly lower than that reported for the P. furiosus .beta.-glucosidase (5.3 vs. 4, 500 s-1 mM-1 with p-nitrophenyl-.beta.-D-glucopyranoside as substrate). The kinetic differences between the two enzymes suggest that, unlike the .beta.-glucosidase, the primary role of the .beta.-mannosidase may not be disaccharide hydrolysis. Other possible roles for this enzyme are discussed.

L9 ANSWER 82 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:584259 CAPLUS
DN 125:266963
TI Cloning, expression, and sequence analysis of the three genes encoding quinoline 2-oxidoreductase, a molybdenum-containing hydroxylase from Pseudomonas putida 86
AU Blase, Marcel; Bruntner, Christina; Tshisuaka, Barbara; Fetzner, Susanne; Lingens, Franz
CS Inst. Mikrobiol. (250), Univ. Hohenheim, Stuttgart, D-70593, Germany
SO Journal of Biological Chemistry (1996), 271(38), 23068-23079 CODEN: JBCHA3; ISSN: 0021-9258
PB American Society for Biochemistry and Molecular Biology
DT Journal
LA English
AB The three genes coding for quinoline 2-oxidoreductase (Qor) or Pseudomonas putida 86 were cloned and sequenced. The qor genes are clustered in the transcriptional order medium (M) small (S), large (L) and code for three subunits of 288 (QorM), 168 (QorS), and 788 (QorL) amino acids, resp. Formation of active quinoline 2-oxidoreductase and degrdn. of quinoline occurred in a recombinant P. putida KT2440 clone. The amino acid sequences of Qor show significant homol. to various prokaryotic molybdenum contg. hydroxylases and to eukaryotic xanthine dehydrogenases. QorS contains two conserved motifs for [2Fe-2S] clusters . The binding motif for the N-terminal [2Fe-2S] cluster corresponds to the binding site of bacterial and chloroplast-type [2Fe-2S] ferredoxins, whereas the amino acid pattern of the internal [2Fe-2S] center apparently is a distinct feature of molybdenum-contg. hydroxylases, showing no homol. to any other described [2Fe-2S] binding motif. The medium subunit QorM presumably contains the FAD, but no conserved sequence areas or described motifs of FAD, NAD, NADP, or ATP binding were detected. Putative binding sites of the molybdopterin cytosine dinucleotide cofactor were detected in QorL by comparison with "contacting segments" recently described in aldehyde oxidoreductase from Desulfovibrio gigas (Romano, M. J., Archer, M., Moura, I., Moura, J. J. G., LeGall, J., Engh, R., Schneider, M., Hof, P., and Huber, R. (1995) Science 270, 1170-1176).

L9 ANSWER 83 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:536734 CAPLUS
DN 125:213780
TI Sequence and transcriptional analysis of the ubiquitin gene cluster in the genome of Spodoptera exigua nucleopolyhedrovirus
AU van Strien, Elisabeth A.; Jansen, Bastiaan J. H.; Mans, Ruud M. W.; Zuidema, Douwe; Vlak, Just M.
CS Dep. Virol., Wageningen Agric. Univ., Wageningen, 6709 PD, Neth.
SO Journal of General Virology (1996), 77(9), 2311-2319 CODEN: JGVIAY; ISSN: 0022-1317
PB Society for General Microbiology
DT Journal
LA English
AB The nucleotide sequence of a 1200 bp DNA fragment of Spodoptera exigua nucleopolyhedrovirus (SeMNPV) was detd. This sequence contained a cluster of two open reading frames (ORFs), one coding for a viral ubiquitin (v-ubi) and another with homol. to orf2 of Autographa californica (Ac) MNPV and Orgyia pseudotsugata (Op) MNPV. The v-ubi ORF is 240 nucleotides (nt) long, potentially encoding a protein of 80 amino acids with a predicted mol. mass of 9.4 kDa. The amino acid sequence of the v-ubi gene in SeMNPV has 75% and 81.6% identity with the v-ubi gene of AcMNPV and OpMNPV and approx. 84% with cellular ubiquitins. Northern blot anal. revealed three major small transcripts late in infection, of about 690, 550 and 400 nt long. Primer extension anal. showed that transcription started from within two consensus late promoter elements (TAAG), located at positions -6 and -30. The start site at position -4/-5 precedes the shortest leader reported to date for a baculovirus gene. The other ORF, xb187, was identified in the opposite orientation immediately upstream of the v-ubi gene. This ORF potentially encodes a 22 kDa protein with unknown function and about 60% amino acid similarity to the products of the orf2 genes of AcMNPV and OpMNPV. The SeMNPV xb187 ORF is transcribed late in infection via two transcripts, 1.2 kb and 770 nt long. The v-ubi-xb187 gene cluster is located at map unit (m.u.) 89 on the genome of SeMNPV. This is different from the position of an identical cluster in the AcMNPV and OpMNPV genomes, located at relative m.u. 20.

L9 ANSWER 84 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:489355 CAPLUS
DN 125:213673
TI Characterization of Newcastle disease virus vaccines by biological properties and sequence analysis of the hemagglutinin-neuraminidase protein gene
AU Seal, Bruce S.; King, Daniel J.; Bennett, Joyce D.
CS Agricultural Research Service, U.S.D.A., Athens, GA, 30605, USA
SO Vaccine (1996), 14(8), 761-766 CODEN: VACCDE; ISSN: 0264-410X
PB Elsevier
DT Journal
LA English
AB Six com. available monovalent Newcastle disease virus (NDV) live-vaccines were examd. for their biol. and genomic stability in comparison to their stated parent virus. Thermostability of the hemagglutinin at 56.degree. for 5 min was consistently obsd. among the majority of the vaccine viruses. One exception was a recently developed NDV vaccine isolated from turkeys that had a thermostability of 15 min.

Neuraminidase activity, as measured by elution rate of agglutinated red blood cells, varied among vaccine viruses and correlated with that of the parent isolate. Virulence as measured by intracerebral pathogenicity index ranged from 0 to 0.39 among NDV vaccine-type viruses, well within the range of avirulent lentogens. Sequence of the fusion protein cleavage site from all the NDV vaccine isolates examd. was consistent with that for lentogens. The entire hemagglutinin-neuraminidase gene sequence was 98% similar among all the NDV vaccine viruses examd. and phylogenetic classification of com. vaccine types correlated with their resp. parent virus. Consequently, the com. produced NDV vaccines reported here appear relatively stable when mass produced in avian embryonated eggs.

L9 ANSWER 85 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:445617 CAPLUS
DN 125:108786
TI Conservation of aconitase residues revealed by multiple sequence analysis . Implications for structure/function relationships
AU Frishman, Dmitrij; Hentze, Matthias W.
CS European Molecular Biology Laboratory, Heidelberg, D-69012, Germany
SO European Journal of Biochemistry (1996), 239(1), 197-200
CODEN: EJBCAI; ISSN: 0014-2956
PB Springer
DT Journal
LA English
AB Aconitases have recently regained much attention, because one member of this family, iron regulatory protein -1 (IRP-1), has been found to play a dual role as a cytoplasmic aconitase and a regulatory RNA-binding protein . This finding has highlighted a novel role for Fe-S clusters as post-translational regulatory switches. We have aligned 28 members of the Fe-S isomerase family, identified highly conserved amino acid residues, and integrated this information with data on the crystallog. structure of mammalian mitochondrial aconitase. We propose structural and/or functional roles for the previously unrecognized conserved residues. Our findings illustrate the value of detailed protein sequence anal. when high-resoln. crystallog. data are already available.

L9 ANSWER 86 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:210807 CAPLUS
DN 124:282377
TI Relationships between bacterial drug resistance pumps and other transport proteins
AU Parish, J. H.; Bentley, J.
CS Dep. Biochem., Univ. Leeds, Leeds, LS2 9JT, UK
SO Journal of Molecular Evolution (1996), 42(2), 281-93
CODEN: JMEVAU; ISSN: 0022-2844
PB Springer
DT Journal
LA English
AB The authors have used three ref. sequences representative of bacterial drug resistance pumps and sugar transport proteins to collect the 91 most closely related sequences from a composite, nonredundant protein sequence database. Having eliminated certain very close relatives, the remainder were subjected to anal. and alignment by using two different similarity matrixes: one of these was a matrix based on structural conservation of amino acid residues in proteins of known conformation and the other was based on the more familiar mutational matrix. Unrooted similarity trees for these proteins were constructed for each matrix and compared. A systematic anal. of the differences between these trees was undertaken and the sequences were analyzed for the presence or absence of certain sequence by the two methods are broadly comparable but that there are some clusters of sequences that are significantly different. Further anal. confirmed that (1) the sequences collected by this objective method are all known or putative 12-helix (in some cases reported as 14-helix) transmembrane proteins , (2) there is evidence for few cases of an origin based on gene duplication, (3) the bacterial drug resistance pumps are distributed in more than one clade and cannot be regarded as a definitive subset of these proteins , and (4) the diversity is such that there is no evidence of a single ancestral protein . The possible extension of the methods to other cases of divergent protein sequences is discussed.

L9 ANSWER 87 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:206985 CAPLUS
DN 124:254534
TI Computational sequence analysis of matrix metalloproteinases
AU Sang, Qingxiang Amy; Douglas, Damon A.
CS Institute Molecular Biophysics, Florida State Univ., Tallahassee, FL, 32306-3006, USA
SO Journal of Protein Chemistry (1996), 15(2), 137-60
CODEN: JPCHD2; ISSN: 0277-8033
PB Plenum
DT Journal
LA English
AB Matrix metalloproteinases (MMP) play a cardinal role in the breakdown of extracellular matrix involved in a variety of biol. and pathol. processes. Research on MMPs has classified and characterized these enzymes according to their matrix substrate specificity, gene and protein domain structure, and regulation of activity and expression. However, the discovery of new MMPs has introduced a need for a more comprehensive and systematic method of classification and quant. comparison of known and newly discovered members. This study compiles a sequence alignment , constructs a dendrogram, and calcs. phys. data and homol. percentage assignments in order to obtain further insight into MMP structure-function relationships. Thorough anal. of MMP primary sequence domains, phys. data patterns, and statistical anal. of sequence homol. yields higher resoln. in the similarities and differences that group MMP members.

L9 ANSWER 88 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:185046 CAPLUS
DN 124:252301
TI An endo-.beta.-1,4-glucanase gene (celA) from the rumen anaerobe Ruminococcus albus 8: cloning, sequencing , and transcriptional analysis
AU Attwood, Graeme T.; Herrera, Felicitas; Weissenstein, Lee A.; White, Bryan A.
CS Dep. Animal Sci., Univ. Illinois Urbana, Urbana, IL, 61801, USA
SO Canadian Journal of Microbiology (1996), 42(3), 267-78
CODEN: CJMIAZ; ISSN: 0008-4166
PB National Research Council of Canada
DT Journal
LA English
AB A genomic library of Ruminococcus albus 8 DNA was constructed in Escherichia coli using bacteriophage .lambda.ZapII. This library was screened for cellulase components and several Ostazin brilliant red/CM-cellulose pos. clones were isolated. All of these clones contained a common 3.4-kb insert, which was recovered as a plasmid by helper phage excision. The carboxymethyl cellulase coding region

was localized to a 1.4-kb region of DNA by nested deletions, and a clone contg. the entire celA gene was sequenced. Anal. of the sequence revealed a 1231-bp open reading frame, coding for a protein of 411 amino acids with a predicted mol. wt. of 45 747. This protein , designated CelA, showed extensive homol. with family 5 endoglucanases by both primary amino acid sequence alignment and hydrophobic cluster anal. Cell-free exts. of E. coli contg. the celA clone demonstrated activity against CM-cellulose and acid swollen cellulose but not against any of the p-nitrophenol glycosides tested, indicating an endo-.beta.-1,4-glucanase type of activity. In vitro transcription-translation expts. showed that three proteins of 48000,44000, and 23000 mol. wt. were produced by clones contg. the celA gene. Northern anal. of RNA extd. from R. albus 8 grown on cellulose indicated a celA transcript of approx. 2700 bases, whereas when R. albus 8 was grown on cellobiose, celA transcripts of approx. 3000 and 600 bases were detected. Primer extension anal. of these RNAs revealed different transcription initiation sites for the celA gene when cells were grown with cellulose or cellobiose as the carbon source. These two sites differed by 370 bases in distance. A model, based on transcription and sequence data, is proposed for celA regulation.

L9 ANSWER 89 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:183892 CAPLUS
DN 124:283626
TI Visualization of protein sequences using the two-dimensional hydrophobic cluster analysis method
AU Semertzidis, Michel T.; Thoreau, Etienne; Tasso, Anne; Henrissat, Bernard; Callebaut, Isabelle; Mornon, Jean Paul
CS Laboratoire de Mineralogie-Cristallographie, Universites Pierre et Marie Curie, Paris, F-75252/05, Fr.
SO Visualizing Biological Information (1995), 129-44. Editor(s): Pickover, Clifford A. Publisher: World Scientific, Singapore, Singapore. CODEN: 62MPAP
DT Conference
LA English
AB This paper describes a method for displaying protein sequences in the form of 2-D graphics, known as the hydrophobic cluster anal. (HCA) method. This technique, essentially visual, makes it possible to compare and align reliably protein sequences that exhibit <20% similarities. HCA may also be of interest in predicting secondary structures from amino acid sequences. The usefulness of the method is demonstrated through a series of examples extd. from the most recent literature. A discussion of the superiority of helical nets over other 2-D plots is included.

L9 ANSWER 90 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:133593 CAPLUS
DN 124:195615
TI Genetic variation in Tula hantaviruses: sequence analysis of the S and M segments of strains from Central Europe
AU Plyusnin, Alexander; Cheng, Ying; Vapalahti, Olli; Pejoch, Milan; Unar, Jiri; Jelinkova, Zuzana; Lehvaeslaiho, Heikki; Lundkvist, Aake; Vaheri, Antti
CS Haartman Inst., Univ. Helsinki, Helsinki, FIN-00014, Finland
SO Virus Research (1995), 39(2-3), 237-50 CODEN: VIREDF; ISSN: 0168-1702
PB Elsevier
DT Journal
LA English
AB Hantavirus carried by the European common vole Microtus arvalis from Moravia (Czech Republic) was analyzed by RT-PCR-sequencing and by reactivity with a panel of monoclonal antibodies (MAbs). Sequencing of the full-length S segment and the proximal part of the M segment showed that the virus belonged to genotype Tula (TUL) we discovered earlier in Microtus arvalis from Central Russia. This finding supported the concept of host dependence of hantaviruses. Phylogenetic analyses suggested a similar evolutionary history for S and M genes of TUL strains; thus far there is no evidence for reassortment in TUL. Geog. clustering of TUL genetic variants was obsd. and different levels of the genetic variability were revealed resembling those estd. for another hantavirus, Puumala (PUU). Comparison of the deduced N protein sequence from Russia and from Moravia showed that genetic drift in TUL occurred not only by accumulation of point mutations but also by the deletion of a nucleotide triplet. It encoded Ser252 which was located within a highly variable hydrophilic part of the N protein carrying B-cell epitopes and presumably forming a loop. Anal. of naturally expressed TUL N-antigen derived from lung tissue of infected voles with MAb indicated antigenic heterogeneity among TUL strains.

L9 ANSWER 91 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1996:69351 CAPLUS
DN 124:136912
TI Taxonomy of simple amino acid sequences
AU Wootton, John C.; Federhen, Scott
CS National Center Biotechnology Information, National Institutes Health, Bethesda, MD, 20894, USA
SO Bioinformatics & Genome Research, Proceedings of the International Conference, 3rd, Tallahassee, June 1-4, 1994 (1995), Meeting Date 1994, 161-73. Editor(s): Lim, Hwa A.; Cantor, Charles R. Publisher: World Scientific, Singapore, Singapore. CODEN: 62FLAC
DT Conference
LA English
AB Approx. one quarter of the amino acids in genome-encoded protein sequences are located in compositionally biased regions of polypeptides . These low-complexity or "simple" sequences contrast with the relatively familiar class of globular protein domains which have quasi-random, high-complexity amino acid compns. The mol. structures, dynamics and interactions of the great majority of low-complexity regions of proteins are unknown. To explore the diversity of these sequences , and to analyze the statistical heterogeneity in the protein sequence databases, the authors have applied math. and computational classification methods. Conventional algorithms for sequence alignment and neighboring (pairwise comparison) fail for low-complexity sequences because of their compositional bias and intricate patterns of variation and evolution. Instead, abstr. sequence properties such as residue and k-gram compn., compositional complexity and repeat periodicity have been used as the basis for statistical clustering , Bayesian classification and neighboring. Multiple Dirichlet densities have been computed to model the statistical heterogeneity of low-complexity sequences. The resulting classification corresponds well to intuitive views of the taxonomy of these regions of proteins as, for example, "glutamine-rich" or "glycine-proline -rich". However, the methods do not capture aspects of the structural, functional and evolutionary diversity of these sequences and new structure-based approaches are also required.

L9 ANSWER 92 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:987474 CAPLUS
DN 124:79836
TI Sequence similarity analysis of Escherichia coli proteins : functional and evolutionary implications
AU Koonin, Eugene V.; Tatusov, Roman L.; Rudd, Kenneth E.

CS Natl. Cent. Biotechnol. Information, Natl. Library Med., Bethesda, MD, 20894, USA
SO Proceedings of the National Academy of Sciences of the United States of America (1995), 92(25), 11921-5 CODEN: PNASA6; ISSN: 0027-8424
PB National Academy of Sciences
DT Journal
LA English
AB A computer anal. of 2328 protein sequences comprising about 60% of the Escherichia coli gene products was performed using methods for database screening with individual sequences and alignment blocks. A high fraction of E. coli proteins - 86% - shows significant sequence similarity to other proteins in current databases; about 70% show conservation at least at the level of distantly related bacteria, and about 40% contain ancient conserved regions (ACRs) shared with eukaryotic or Archael proteins . For >90% of the e. coli proteins , either functional information or sequence similarity, or both, are available. Forty-six percent of the E. coli proteins belong to 299 clusters of paralogs (intraspecies homologs) defined on the basis of pairwise similarity. Another 10% could be included in 70 superclusters contain only two to four members. In contrast, nearly 25% of all E. coli proteins belong to the four largest superclusters - namely, permeases, ATPases and GTPases with the conserved "Walker-type" motif, helix-turn-helix regulatory proteins , and NAD(FAD)-binding proteins . We conclude that bacterial protein sequences generally are highly conserved in evolution, with about 50% of all ACR-contg. protein families represented among the E. coli gene products. With the current sequence databases and methods of their screening, computer anal. yields useful information on the functions and evolutionary relationships of the vast majority of genes in a bacterial genome. Sequence similarity with E. coli proteins allows the prediction of functions for a no. of important eukaryotic genes, including several whose products are implicated in human diseases.

L9 ANSWER 93 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:841524 CAPLUS
DN 124:2108
TI Cloning, sequencing , and phenotypic analysis of laf1, encoding the flagellin of the lateral flagella of Azospirillum brasilense Sp7
AU Moens, Sara; Michiels, Kris; Keijers, Veerle; Van Leuven, Fred; Vanderleyden, Jos
CS F. A. Janssens Lab. Genet., Katholieke Univ. Leuven, Louvain, B-3000, Belg.
SO Journal of Bacteriology (1995), 177(19), 5419-26 CODEN: JOBAAY; ISSN: 0021-9193
PB American Society for Microbiology
DT Journal
LA English
AB A. brasilense can display a single polar flagellum and several lateral flagella. The A. brasilense Sp7 gene laf1, encoding the flagellin of the lateral flagella, was isolated and sequenced. The derived protein sequence is extensively similar to those of the flagellins of Rhizobium meliloti, Agrobacterium tumefaciens, Bartonella bacilliformis, and Caulobacter crescentus. An amino acid alignment shows that the flagellins of these bacteria are clustered and are clearly different from other known flagellins. A laf1 mutant, FAJ0201, was constructed by replacing an internal part of the laf1 gene by a kanamycin resistance-encoding gene cassette. The mutant is devoid of lateral flagella but still forms the polar flagellum. This phenotype is further characterized by the abolishment of the capacities to swarm on a semisolid surface and to spread from a stab inoculation in a semisolid medium. FAJ0201 shows a normal wheat root colonization pattern in the initial stage of plant root interaction.

L9 ANSWER 94 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:785745 CAPLUS
DN 123:192975
TI Characterization of protein structure/function relationship by sequence analysis without previous alignment : distinction between sub-groups of protein kinases
AU Guerrucci, Marie-Anne; Belle, Robert
CS Atelier de Bioinformatique, Institut Curie, Paris, 75007, Fr.
SO Bioscience Reports (1995), 15(3), 161-71 CODEN: BRPTDT; ISSN: 0144-8463
PB Plenum
DT Journal
LA English
AB Using an approach for protein comparison by computer anal. based on signal treatment methods without previous alignment of the sequence , the authors have analyzed the structure/function relation of related proteins . The aim was to demonstrate that from a few members of related proteins , specific parameters can be obtained and used for the characterization of newly sequenced proteins obtained by mol. biol. techniques. The anal. was performed on protein kinases, which comprise the largest known family of proteins , and therefore allows valid estns. to be made. The authors show that using only a dozen defined proteins , the specific parameters extd. from their sequences classified the protein kinase family into two sub-groups: the protein serine/threonine kinases (PSKs) and the protein tyrosine kinases (PTKs). The anal., largely involving computation, appears applicable to large scale data-bank anal. and prediction of protein functions.

L9 ANSWER 95 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:744501 CAPLUS
DN 123:136411
TI Recurring local sequence motifs in proteins
AU Han, Karen F.; Baker, David
CS Grad. Group Biophys., Univ. California, San Francisco, CA, 94143, USA
SO Journal of Molecular Biology (1995), 251(1), 176-87 CODEN: JMOBAK; ISSN: 0022-2836
PB Academic
DT Journal
LA English
AB We describe a completely automated approach to identifying local sequence motifs that transcend protein family boundaries. Cluster anal. is used to identify recurring patterns of variation at single positions and in short segments of contiguous positions in multiple sequence alignments for a non-redundant set of protein families. Parallel expts. on simulated data sets constructed with the overall residue frequencies of proteins but not the inter-residue correlations show that naturally occurring protein sequences are significantly more clustered than the corresponding random sequences for window lengths ranging from one to 13 contiguous positions. The patterns of variation at single positions are not in general surprising: chem. similar amino acids tend to be grouped together. More interesting patterns emerge as the window length increases. The patterns of variation for longer window lengths are in part recognizable patterns of hydrophobic and hydrophilic residues, and in part less obvious combinations. A particularly interesting class of patterns features highly conserved glycine residues. The patterns provide a means to abstr. the information contained in multiple sequence alignments and may be useful for

comparison of distantly related sequences or sequence families and for protein structure prediction.

L9 ANSWER 96 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:715576 CAPLUS
DN 123:249537
TI The PTR family: a new group of peptide transporters
AU Steiner, Henry-York; Naider, Fred; Becker, Jeffrey M.
CS Dep. Microbiology, Univ. Tennessee, Knoxville, TN, 37996-0845, USA
SO Molecular Microbiology (1995), 16(5), 825-34 CODEN: MOMIEE; ISSN: 0950-382X
PB Blackwell
DT Journal; General Review
LA English
AB The transport of peptides into cells is a well-documented biol. phenomenon which is accomplished by specific, energy-dependent transporters found in a no. of organisms as diverse as bacteria and humans. Until recently, the majority of peptide transporters cloned and characterized were found to be proteins of the ATP-binding cassette (ABC) family. A new family of peptide transporters is called the PTR family. This group of proteins , distinct from the ABC-type peptide transporters, was uncovered by sequence analyses of a no. of recently discovered peptide transport proteins . Alignment of these proteins demonstrated a high no. of identical and similar residues and identified conserved glycosylation and phosphorylation sites, as well as a structural motif unique to this group of proteins . Cluster anal. among the proteins indicated these sequences were indeed related and could be further divided into 2 subfamilies. A phylogenetic anal. of these new peptide transport sequences, compared to over 50 other peptide and membrane-bound transporters, showed that these proteins comprise a distinct, sep. group of proteins .

L9 ANSWER 97 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:694641 CAPLUS
DN 123:307650
TI Cloning and study of the genetic organization of the exe gene cluster of Aeromonas salmonicida
AU Karlyshev, Andrey V.; MacIntyre, Sheila
CS Dep. Microbiol., Univ. Reading, Whiteknights, Reading, Berkshire, UK
SO Gene (1995), 158(1), 77-82 CODEN: GENED6; ISSN: 0378-1119
PB Elsevier
DT Journal
LA English
AB The Aeromonas salmonicida (As) exe gene cluster , an addnl. member of the pul-related operon family required for general signal-sequence-dependent secretion of proteins from Gram- bacteria, was cloned in the broad-host-range cosmid pLAFR3. Twelve genes, exeC-N, were identified by partial nucleotide (nt) sequence analyses (exeE-N) or detn. of the complete sequence (exeC and exeD). The organization of the exeC-N genes is similar to that of several other operons of this family. These genes are arranged contiguously and are apparently transcribed in the same direction. On alignment of As and A. hydrophila exe sequences a 73-bp 'silent' deletion was identified close to the end of the As exeF gene. No gene encoding prepilin peptidase (the PulO homolog) was detected in this region. The exeN gene is evidently the last gene of this operon; it is followed by an ORF encoding a putative transcription regulator.

L9 ANSWER 98 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:479397 CAPLUS
DN 123:48983
TI Sequence polymorphism in the 5'NTR and in the P1 coding region of potato virus Y genomic RNA
AU Tordo, V. Marie-Jeanne; Chachulska, A. M.; Fakhfakh, H.; Le Romancer, M.; Robaglia, C.; Astier-Manifacier, S.
CS Lab. Pathol. Veg., INRA, Versailles, 78026, Fr.
SO Journal of General Virology (1995), 76(4), 939-49 CODEN: JGVIAY; ISSN: 0022-1317
PB Society for General Microbiology
DT Journal
LA English
AB Potato virus Y (PVY) the type member of the genus Potyvirus, occurs world-wide as isolates which differ in host range and the type of symptoms caused. The sequences of a 5' segment of viral RNA overlapping the 5' non-translated region (5'NTR) alone (10 isolates) or the 5'NTR and the adjacent P1 coding region (8 isolates) were established. These data were used to quantify the polymorphism in the 5'-terminal part of the PVY genome. Nucleotide sequence identity between isolates ranged from 66-100% in the 5'NTR and from 70-100% in the P1 coding region. The lowest amino acid sequence similarity between PVY P1 was 77%, illustrating the high variability of this protein in the PVY species. Phylogenetic trees based on either 5'NTR or P1 sequences analyses resulted in the same clustering of the studied isolates into 3 groups. Group I comprises potato isolates all inducing tobacco veinal necrosis symptoms. Group II contains isolates inducing either tobacco veinal necrosis or mosaic symptoms in tobacco. Group III contains mainly pepper or tomato isolates inducing mosaic symptoms in tobacco and shows a geog. clustering of the Tunisian isolates. This clustering into 3 groups is discussed in comparison with phylogenetic trees previously obtained from capsid gene or 3'NTR sequence anal. in the PVY species. Multiple sequence alignment indicated conserved motifs potentially involved in viral functions.

L9 ANSWER 99 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:421841 CAPLUS
DN 123:26898
TI Nucleotide sequence and transcriptional analysis of the celD .beta.-glucanase gene from Ruminococcus flavefaciens FD-1
AU Vercoe, Philip E.; Spight, Donn H.; White, Bryan A.
CS Dep. of Animal Sciences, Univ. of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA
SO Canadian Journal of Microbiology (1995), 41(1), 27-34 CODEN: CJMIAZ; ISSN: 0008-4166
PB National Research Council of Canada
DT Journal
LA English
AB The nucleotide sequence of the celD gene, which encodes endoglucanase and xylanase activity, from Ruminococcus flavefaciens FD-1 was detd. The DNA sequence of celD contains an open reading frame of 1215 nucleotides that encodes a polypeptide of 405 amino acids with a mol. mass of 44,631 Da. The primary amino acid sequence of CelD was screened against the GenBank data base for similar polypeptide sequences and the anal. indicated that CelD has common features with endoglucanases from the family E cellulases. Both hydrophobic cluster and BESFIT (Genetics Computer Group (University of Wisconsin) package) analyses confirmed this relationship. Pairwise alignments using BESTFIT revealed that CelD was most closely related to endE4 from Thermomonospora fusca over a 160 amino acid window. The histidine, aspartate, and glutamate residues identified as being essential for catalytic activity in family E cellulases are conserved in CelD. A Shine-Dalgarno-like sequence was

present 5 base pairs (bp) upstream of the translation start site. Primer extension anal. indicated that different transcription initiation sites are used to initiate transcription of CelD in Escherichia coli and R. flavefaciens. In the case of R. flavefaciens the transcription initiation site is at a T residue (nucleotide 273) 16 bp upstream from the translational start site. A region resembling a .sigma.70-like-10 promoter sequence is present upstream from the transcription initiation site, but there is no apparent - 35 region. In contrast, transcription in E. coli is initiated at a C residue 258 bp upstream from the translational start site and a sequence resembling a .sigma.70-like-10 region is present 5 bp upstream of this residue. Assuming 17 bp is the optimal distance between -10 and -35 sites for .sigma.70 consensus sequences, the -35 region for CelD transcription initiation in E. coli would be outside the boundaries of the cloned R. flavefaciens DNA.

L9 ANSWER 100 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:215168 CAPLUS
DN 122:307656
TI Characterization and sequence analysis of the lsg (LOS synthesis genes) locus from Haemophilus influenzae type b
AU McLaughlin, R.; Lee, N.-G.; Abu Kwaik, Y.; Spinola, S. M.; Apicella, M. A.
CS Health Sciences Center, University of Oklahoma, OK, 73190, USA
SO Journal of Endotoxin Research (1994), 1(3), 165-74
CODEN: JENREB; ISSN: 0968-0519
DT Journal
LA English
AB Anal. of the lsg (LOS synthesis genes) cluster in Escherichia coli strain K12 and mutations in the lsg locus in Haemophilus influenzae type b indicated the presence of 3 regions responsible for sequential modifications of E. coli lipopolysaccharide (LPS). Sequencing of the lsg region yielded 7,435 bp that encompassed 7 complete and 1 partial open reading frames (ORFs 1-8). The predicted product of ORF1 had homol. to the consensus sequence of cytochrome b proteins (21% identity, 51% similarity) and to other transmembrane proteins . The products of ORF5 and ORF6 share overall 23% identity and 49% similarity with each other. The ORF6 protein had high homol. with the product of ORF275 of the E. coli rfb gene cluster (40% identity, 58% similarity), whose function is not known. Multiple sequence alignment of the ORF5 and ORF6 proteins with the RfbB, RfbJ and RfbX proteins revealed conserved motifs over the N-terminal half region of all these proteins . The products of ORF7 and ORF8 are homologous with Azotobacter vinelandii MolA protein (30% identity, 51% similarity) and MolB protein (26% identity, 48% similarity), resp. The promoter regions of ORF1, 7 and 8 were detd. by primer extension anal. and similar to bacterial .sigma.70-dependent promoters. ORF7 and ORF8 are transcribed into diverse orientation. At least 5 of the encoded proteins have been identified using coupled E. coli transcription/translation system and labeling with [35S]-methionine. The authors conclude that the genetic organization of the lsg biosynthesis pathway involves multiple operons that lead to the assembly of an H. influenzae LOS structure.

L9 ANSWER 101 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:213255 CAPLUS
DN 122:48144
TI Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in Staphylococcus aureus
AU Lin, Wen S.; Cunneen, Tim; Lee, Chia Y.
CS Dep. Micribiol., Univ. Kansas Med. Cent., Kansas City, KS, 66160, USA
SO Journal of Bacteriology (1994), 176(22), 7005-16 CODEN: JOBAAY; ISSN: 0021-9193
PB American Society for Microbiology
DT Journal
LA English
AB A 19.4-kb DNA region contg. a cluster of genes affecting type 1 capsule prodn. was previously cloned from Staphylococcus aureus M. Subcloning expts. showed that these capsule (cap) genes are localized in a 14.6-kb region. Sequencing anal. of the 14.6-kb fragment revealed 13 open reading frames (ORFs). Complementation tests were used to map a collection of Cap- mutations in 10 of the 13 ORFs, indicating that these 10 genes are involved in capsule biosynthesis. The requirement for the remaining three ORFs in the synthesis of the capsule was demonstrated by constructing site-specific mutations corresponding to each of the three ORFs. An Escherichia coli S30 in vitro transcription-translation system clearly identified 7 of the 13 proteins predicted from the ORFs. Homol. search between the predicted proteins and those in the data bank showed very high homol. (52.3% identity) between capL and vipA, moderate homol. (29% identity) between capI and vipB, and limited homol. (21.8% identity) between capM and vipC. The vipA, vipB, and vipC genes were shown to be involved in the biosynthesis of Salmonella typhi Vi antigen, a homopolymer polysaccharide consisting of N-acetylgalactosamino uronic acid, which is also one of the components of the staphylococcal type 1 capsule. The homol. between these sets of genes therefore suggests that capL, capI, and capM may be involved in the biosynthesis of amino sugar, N-acetylgalactosamino uronic acid. In addn., the search showed that CapG aligned well with the consensus sequence of a family of acetyltransferases from various prokaryotic organisms, suggesting that CapG may be an acetyltransferase. Using the isogenic Cap- and Cap+ strains constructed in this study, it was confirmed that type 1 capsule is an important virulence factor in a mouse lethality test.

L9 ANSWER 102 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:199279 CAPLUS
DN 122:25240
TI Phylogenetic relationships reveal recombination among isolates of cauliflower mosaic virus
AU Chenault, Kelly D.; Melcher, Ulrich
CS Dep. Biochem. Molecular Biol., Oklahoma State Univ., Stillwater, OK, 74078, USA
SO Journal of Molecular Evolution (1994), 39(5), 496-505
CODEN: JMEVAU; ISSN: 0022-2844
PB Springer
DT Journal
LA English
AB Isolates of cauliflower mosaic virus (CaMV) differ in host range and symptomatol. Knowledge of their sequence relationships should assist in identifying nucleotide sequences responsible for isolate-specific characters. Complete nucleotide sequences of the DNAs of 8 isolates of CaMV were aligned and the aligned sequences were used to analyze phylogenetic relationships by max. likelihood, bootstrapped parsimony, and distance methods. Isolates found in North America clustered sep. from those isolated from other parts of the world. Addnl. isolates, for which partial sequences were available, were incorporated into phylogenetic anal. of the sequences of genome segments corresponding to individual protein coding regions or the large intergenic region of CaMV DNA. The anal.

revealed several instances where the position of an isolate on a tree for one coding region did not agree with the position of the isolate on the tree for the complete genome or with its position on trees for other coding regions. Examn. of the distribution of shared residue types of phylogenetically informative positions in anomalous regions suggested that most of the anomalies were due to recombination events during the evolution of the isolates. Application of an algorithm that searches for segments of significant length that are identical between pairs of isolates or contain a significantly high concn. of polymorphisms suggested two addnl. recombination events between progenitors of the isolates studied and an event between the XinJing isolate and a CaMV not represented in the data set. An earlier phylogenetic origin for CaMV than for carnation etched ring virus, the caulimovirus used as outgroup in these analyses, was deduced from the position of the outgroup with North American isolates in some trees, but with non-North American isolates in other trees.

L9 ANSWER 103 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:173446 CAPLUS
DN 122:152891
TI Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences
AU Korber, Bette T. M.; Kunstman, Kevin J.; Patterson, Bruce K.; Furtado, Manohar; McEvilly, Miranda M.; Levy, Robert; Wolinsky, Steven M.
CS Theoretical Biology and Biophysics (T10), Los Alamos National Laboratory, Los Alamos, NM, 87545, USA
SO Journal of Virology (1994), 68(11), 7467-81 CODEN: JOVIAM; ISSN: 0022-538X
PB American Society for Microbiology
DT Journal
LA English
AB Human immunodeficiency virus type 1 (HIV-1) sequences were generated from blood and from brain tissue obtained by stereotactic biopsy from 6 patients undergoing a diagnostic neurosurgical procedure. Proviral DNA was directly amplified by nested PCR, and 8-36 clones from each sample were sequenced. Phylogenetic anal. of intrapatient envelope V3-V5 region HIV-1 DNA sequence sets revealed that brain viral sequences were clustered relative to the blood viral sequences, suggestive of tissue-specific compartmentalization of the virus in 4 of the 6 cases. In the other 2 cases, the blood and brain virus sequences were intermingled in the phylogenetic analyses , suggesting trafficking of virus between the 2 tissues. Slide-based PCR-driven in situ hybridization of 2 of the patients' brain biopsy samples confirmed this interpretation of the intrapatient phylogenetic analyses. Interpatient V3 region brain-derived sequence distances were significantly less than blood-derived sequence distances. Relative to the tip of the loop, the set of brain-derived viral sequences had a tendency towards neg. or neutral charge compared with the set of blood-derived viral sequences. Entropy calcns. were used as a measure of the variability at each position in alignments of blood and brain viral sequences. A relatively conserved set of positions were found, with a significantly lower entropy in the brain- than in the blood-derived viral sequences. These sites constitute a brain "signature pattern," or a non-contiguous set of amino acids in the V3 region conserved in viral sequences derived from brain tissue. This brain-derived signature pattern was also well preserved among isolates previously characterized in vitro as macrophage tropic. Macrophage-monocyte tropism may be the biol. constraint that results in the conservation of the viral brain signature pattern.

L9 ANSWER 104 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1995:29769 CAPLUS
DN 122:74871
TI A sequence analysis of lipases, esterases and related proteins
AU Petersen, Steffen B.; Drablos, Finn
CS Natural Science Section, MR Center, Trondheim, 7034, Norway
SO Lipases (1994), 23-48, 1 plate. Editor(s): Woolley, Paul; Petersen, Steffen B. Publisher: Cambridge Univ. Press, Cambridge, UK. CODEN: 60HHAW
DT Conference
LA English
AB The search is described for common sequence motifs in a large no. of lipases, esterases and related proteins using MULTIM, a program suite for semi-automatic multiple sequence alignment in protein engineering and protein sequence studies. With few exceptions, all the sequences contained the GxSxG motif, where x is any amino acid residue. A classification of the contexts of the putative active serine showed that the proteins can be grouped into 2 major classes, one displaying the AGY and the other the TCN codon for the active-site serine.

L9 ANSWER 105 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1994:550119 CAPLUS
DN 121:150119
TI The pyrimidine biosynthesis operon of the thermophile Bacillus caldolyticus includes genes for uracil phosphoribosyltransferase and uracil permease
AU Ghim, Sa-Youl; Neuhard, Jan
CS Inst. Molecular Biology, Univ. Copenhagen, Copenhage K, DK-1307, Den.
SO Journal of Bacteriology (1994), 176(12), 3698-707 CODEN: JOBAAY; ISSN: 0021-9193
DT Journal
LA English
AB A 3-kb DNA segment of the Bacillus caldolyticus genome including the 5' end of the pyr cluster has been cloned and sequenced. The sequence revealed the presence of two open reading frames, pyrR and pyrP, located immediately upstream of the previously sequenced pyrB gene encoding the pyrimidine biosynthesis enzyme aspartate transcarbamoylase. The pyrR and pyrP genes encoded polypeptides with calcd. mol. masses of 19.9 and 45.2 kDa, resp. Expression of these ORFs was confirmed by anal. of plasmid-encoded polypeptides in minicells. Sequence alignment and complementation analyses identified the pyrR gene product as a uracil phosphoribosyltransferase and the pyrP gene product as a membrane-bound uracil permease. By using promoter expression vectors, a 650-bp EcoRI-HincII fragment, including the 5' end of pyrR and its upstream region, was found to contain the pyr operon promoter. The transcriptional start point was located by primer extension at a position 153 bp upstream of the pyrR translation initiation codon, 7 bp 3' of a sequence resembling a .sigma.A-dependent Bacillus subtilis promoter. This established the following organization of the ten cistrons within the pyr operon: promoter-pyrR-pyrP-pyrB-pyrC-pyrAa-pyrAb- orf2-pyrD-pyrF-pyrE. The nucleotide sequences of the region upstream of pyrR and of the pyrR-pyrP and pyrP-pyrB intercistronic regions indicated that the transcript may form two mutually exclusive secondary structures within each of these regions. One of these structures resembled a rho-independent transcriptional

terminator. The possible implication of these structures for pyrimidine regulation of the operon is discussed.

L9 ANSWER 106 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1994:239921 CAPLUS
DN 120:239921
TI Molecular evolution of herpesviruses: genomic and protein sequence comparisons
AU Karlin, Samuel; Mocarski, Edward S.; Schachtel, Gabriel A.
CS Dep. Math., Stanford Univ., Stanford, CA, 94305, USA
SO Journal of Virology (1994), 68(3), 1886-902 CODEN: JOVIAM; ISSN: 0022-538X
DT Journal
LA English
AB Phylogenetic reconstruction of herpesvirus evolution is generally founded on amino acid sequence comparisons of specific proteins . These are relevant to the evolution of the specific gene (or set of genes), but the resulting phylogeny may vary depending on the particular sequence chosen for anal. (or comparison). The first part of this report compares 13 herpesvirus genomes by using a new multidimensional methodol. based on distance measures and partial orderings of dinucleotide relative abundances. The sequences were analyzed with respect to (1) genomic compositional extremes; (2) total distances within and between genomes; (3) partial orderings among genomes relative to a set of sequence stds.; (4) concordance correlations of genome distances; and (5) consistency with the alpha-, beta-, gammaherpesvirus classification . Distance assessments within individual herpesvirus genomes show each to be quite homogeneous relative to the comparisons between genomes. The gammaherpesviruses, Epstein-Barr virus (EBV), herpesvirus saimiri, and bovine herpesvirus 4 are both diverse and sep. from other herpesvirus classes, whereas alpha- and betaherpesviruses overlap. The anal. revealed that the most central genome (closest to a consensus herpesvirus genome and most individual herpesvirus sequences of different classes) is that of human herpesvirus 6, suggesting that this genome is closest to a progenitor herpesvirus. The shorter DNA distances among alphaherpesviruses supports the hypothesis that the alpha class is of relatively recent ancestry. Equine herpesvirus 1 (EHV1) stands out as the most central alphaherpesvirus, suggesting that it may approx. an ancestral alphaherpesvirus. Among all herpesviruses, the EBV genome is closest to human sequences. In the DNA partial orderings, the chicken sequence collection is invariably as close as or closer to all herpesvirus sequences than the human sequence collection is, which may imply that the chicken (or other avian species) is a more natural or more ancient host of herpesviruses. In the 2nd part of this report, evolutionary relations among the 13 herpesvirus genomes are evaluated on the basis of recent methods of amino acid alignment applied to 4 essential protein sequences. In this anal., the alignment of the 2 betaherpesviruses (human cytomegalovirus vs. human herpesvirus 6) showed lower scores compared with alignments within alphaherpesviruses (i.e., among EHV1, herpes simplex virus type 1, varicella-zoster virus, pseudorabies virus type 1, and Marek's disease virus) and within gammaherpesviruses (EBV vs. herpesvirus saimiri). Comparisons within the alpha class generally produced the highest alignment scores, with EHV1 and pseudorabies type 1 prominent, whereas herpes simplex virus type 1 vs. varicella-zoster virus show the least similarity among the alpha sequences. The within-alpha, beta, and gamma class sequence similarity scores are generally 50-100% higher than the between-class sequence similarity scores. These results suggest that the betaherpesviruses sepd. earlier than the formation of the gamma class and that the alpha class may be of the most recent ancestry. By these methods, evolutionary relations derived from genomic comparisons vs. protein comparisons differ to some extent. The dinucleotide relative abundance distances appear to discriminate DNA structure specificity more than sequence specificity. The evolutionary development of genes among viruses (and species) is more dependent on each individual gene.

L9 ANSWER 107 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1994:237238 CAPLUS
DN 120:237238
TI The biological properties of a distinct tospovirus and sequence analysis of its S RNA
AU Pang, Sheng Zhi; Slightom, Jerry L.; Gonsalves, Dennis
CS Dep. Plant Pathol., Cornell Univ., Geneva, NY, 14456, USA
SO Phytopathology (1993), 83(7), 728-33 CODEN: PHYTAJ; ISSN: 0031-949X
DT Journal
LA English
AB A tospovirus isolate from Brazil, designated TSWV-B, was first identified as a unique isolate based on the authors' observation that transgenic plants expressing the N gene of the lettuce strain of tomato spotted wilt virus (TSWV-BL) were susceptible to TSWV-B but showed resistance to both TSWV (L type) and impatiens necrotic spot virus (INSV). TSWV-B was serol. distinct from TSWV and INSV. TSWV-B generally incited symptoms resembling those caused by other TSWV isolates, except TSWV-B systemically infected Petunia hybrida, which is a local-lesion host of TSWV. Unlike the cucurbit isolate TSWV-W, TSWV-B did not infect Cucumis sativus and only occasionally induced systemic infections on C. metuliferus. The complete nucleotide sequence of the S RNA of TSWV-B was detd. with cDNA clones to be 3,049 nucleotides long. The genome organization of this S RNA was similar to those of TSWV and INSV. The alignment of the S RNA nucleotide and deduced amino acid sequences with the homologous sequences of TSWV (isolates CNPH1, L3, and BL) and INSV revealed that TSWV-B was related more closely to all the TSWV isolates than to INSV. There was a higher degree of identity among the TSWV isolates than with TSWV-B. Thus, TSWV-B appears to be a distinct tospovirus; however, a precise classification requires addnl. biol. and mol. information on this isolate as well as comparison to other tospovirus isolates.

L9 ANSWER 108 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1994:71955 CAPLUS
DN 120:71955
TI P-type ATPases of eukaryotes and bacteria: sequence analyses and construction of phylogenetic trees
AU Fagan, Matthew J.; Saier, Milton H., Jr.
CS Dep. Biol., Univ. California, San Diego, La Jolla, CA, 92093-0116, USA
SO Journal of Molecular Evolution (1994), 38(1), 57-99 CODEN: JMEVAU; ISSN: 0022-2844
DT Journal
LA English
AB The amino acid sequences of 47 P-type ATPases from several eukaryotic and bacterial kingdoms were divided into three structural segments based on individual hydropathy profiles. Each homologous segment was (1) multiply aligned and functionally evaluated, (2) statistically analyzed to det. the degree of sequence similarity, and (3) used for the construction of parsimonious phylogenetic trees. All of the P-type ATPases analyzed comprise a single family with four major clusters correlating with their cation specificities and

biol. sources as follows: cluster 1: Ca2+ -transporting ATPases; cluster 2: Na+- and gastric H+-ATPases; cluster 3: plasma membrane H+-translocating ATPases of plants, fungi, and lower eukaryotes; and cluster 4: all but one of the bacterial P-type ATPases (specific for K+, Cd2+, Cu2+ and an unknown cation). The one bacterial exception to this general pattern was the Mg2+-ATPase of Salmonella typhimurium, which clustered with the eukaryotic sequences. Although exceptions were noted, the similarities of the phylogenetic trees derived from the the three segments analyzed led to the probability that the N-terminal segments 1 and the centrally localized segments 2 evolved from a single primordial ATPase which existed prior to the divergence of eukaryotes from prokaryotes. By contrast, the C-terminal segments 3 appear to be eukaryotic specific, are not found in similar form in any of the prokaryotic enzymes, and are not all demonstrably homologous among the eukaryotic enzymes. These C-terminal domains may therefore have either arisen after the divergence of eukaryotes from prokaryotes or exhibited more rapid sequence divergence than either segment 1 or 2, thus masking their common origin. The relative rates of evolutionary divergence for the three segments were detd. to be segment 2 < segment 1 < segment 3. Correlative functional analyses of the most conserved regions of these ATPases, based on published site-specific mutagenesis data, provided preliminary evidence for their functional roles in the transport mechanism. They should provide a guide for the design of future studies of structure-function relationships employing mol. genetic, biochem., and biophys. techniques.

L9 ANSWER 109 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1994:24560 CAPLUS
DN 120:24560
TI Molecular cloning and characterization of a human carboxylesterase gene
AU Shibata, Futoshi; Takagi, Yasumitsu; Kitajima, Masato; Kuroda, Toshihisa; Omura, Tsuneo
CS Sch. Med., Fujita Health Univ., Toyoake, 470-11, Japan
SO Genomics (1993), 17(1), 76-82 CODEN: GNMCEP; ISSN: 0888-7543
DT Journal
LA English
AB A cDNA encoding human liver carboxylesterase and its gene were isolated. Nucleotide sequence analyses of the cDNA revealed that the predicted enzyme protein consists of 567 amino acids, including 18 amino acids of a putative signal peptide . Comparison of the deduced amino acid sequences of this enzyme with those of seven other carboxylesterases in various mammalian species, together with exptl. data from several other labs., showed that these enzymes can be classified into three groups depending on the sequences at their carboxyl terminals and the presence or absence of one exon. A human carboxylesterase gene was found to span approx. 30 kb and to have 14 small exons. Alignments of this gene with those of human cholinesterase and rat cholesterol esterase indicated insertional sites at some introns and homologous amino acid sequences around them, although these genes have different nos. of exons. Thus the results supported the conclusion that these esterases evolved from a common ancestral gene.

L9 ANSWER 110 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1994:6409 CAPLUS
DN 120:6409
TI Self- peptides from four HLA-DR alleles share hydrophobic anchor residues near the amino-terminal including proline as a stop signal for trimming
AU Kropchofer, Harald; Max, Heiner; Halder, Thomas; Kalbus, Matthias; Muller, Claudia A.; Kalbacher, Hubert
CS Cent. Med. Natl. Sci., Univ. Tuebingen, Tuebingen, W-7400, Germany
SO Journal of Immunology (1993), 151(9), 4732-42 CODEN: JOIMA3; ISSN: 0022-1767
DT Journal
LA English
AB Naturally processed MHC class II-assocd. peptides proved to be heterogeneous in size, varying from 13 to 25 amino acids. Truncation variants suggested sequence motifs that afford the amino termini to be shifted for obtaining an alignment : a 9-to 11-residue core region that is bordered by primary anchor residues is surrounded by extra sequences of variable lengths and hitherto unknown functions. Herein the authors present bulk sequencing analyses of self- peptides from four HLA-DR alleles and HLA-DQw7 clearly showing that the length of most of the NH2-terminal preanchor sequence is limited to 1 to 3 residues. Most strikingly, proline is the dominant residue reappearing at positions 2 and 3 in any allele. Proline was revealed to function as a stop signal for NH2-terminal trimming as well as a secondary anchor: crude cytosolic and endosomal peptide fractions could be processed by amino-peptidases in vitro, whereupon DR1 binding peptides with increased affinity were generated. In addn., aminopeptidase treatment of DR1:self- peptide complexes implied that proline together with steric constraints of the MHC mol. do protect the peptides ' NH2-termini from further processing, whereas their COOH-termini were accessible to cathepsin B processing. Finally, bulk sequencing profiles contained signals from further putative anchor residues clustering in the NH2-terminal region: tyrosine, phenylalanine, leucine, isoleucine, and valine are enriched at positions 2 to 4 in DR1, DR5 and DR6, however, at positions 4 to 6 in DR3. Isotype-specificity is demonstrated by DQw7 displaying glutamine and asparagine at position 2. Obviously, the degenerate occurrence of arom. or aliph. side chains close to the NH2-terminal guarantees for essential interactions with a hydrophobic pocket of the investigated DR mols. Most probably, this pocket is located in the nonpolymorphic DR .alpha.-chain rationalizing previous findings of promiscuous peptide binding to different DR alleles.

L9 ANSWER 111 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1993:663897 CAPLUS
DN 119:263897
TI Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics
AU Kumazawa, Yoshinori; Nishida, Mutsumi
CS Dep. Mol. Cell Biol., Univ. California, Berkeley, CA, 94720, USA
SO Journal of Molecular Evolution (1993), 37(4), 380-98 CODEN: JMEVAU; ISSN: 0022-2844
DT Journal
LA English
AB Mitochondrial DNA sequences are often used to construct rmol. phylogenetic trees among closely related animals. In order to examine the usefulness of mtDNA sequences for deep-branch phylogenetics, genes in previously reported mtDNA sequences were analyzed among several animals that diverged 20-600 million years ago. Unambiguous alignment was achieved for stem-forming regions of mitochondrial tRNA genes by virtue of their conservative secondary structures. Sequences derived from stem parts of the mitochondrial tRNA genes appeared to accumulate much variation linearly for a long period of time: nearly 100 Myr for transition differences and more than 350 Myr for transversion differences. This

characteristic could be attributed, in part, to the structural variability of mitochondrial tRNAs, which have fewer restrictions on their tertiary structure than do nonmitochondrial tRNAs. The tRNA sequence data served to reconstruct a well-established phylogeny of the animals with 100% bootstrap probabilities by both max. parsimony and neighbor-joining methods. By contrast, mitochondrial protein genes coding for cytochrome b and cytochrome oxidase subunit I did not reconstruct the established phylogeny or did so only weakly, although a variety of fractions of the protein gene sequences were subjected to tree-building. This discouraging phylogenetic performance of mitochondrial protein genes, esp. with respect to branches originating over 300 Myr ago, was not simply due to high randomness in the data. It may have been due to the relative susceptibility of the protein genes to natural selection as compared with the stem parts of mitochondrial tRNA genes. Thus, mitochondrial tRNA genes may be useful in resolving deep branches in animal phylogenies with divergences that occurred some hundreds of Myr ago. For this purpose, the authors designed a set of primers with which mtDNA fragments encompassing clustered tRNA genes were successfully amplified from various vertebrates by the polymerase chain reaction.

L9 ANSWER 112 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1993:512844 CAPLUS
DN 119:112844
TI Dot-plot comparisons by multivariate analysis (DOCMA): A tool for classifying protein sequences
AU Landes, Claudine; Henaut, Alain; Risler, Jean Loup
CS Cent. Genet. Mol., Univ. Pierre et Marie Curie, Gif-sur-Yvette, 91198, Fr.
SO CABIOS, Computer Applications in the Biosciences (1993), 9(2), 191-6 CODEN: COABER; ISSN: 0266-7061
DT Journal
LA English
AB A method aimed at classifying protein sequences without resorting to pairwise alignment is presented. Called DOCMA (DOt-plot Comparisons by Multivariate Anal.), it is based on a multivariate anal. of the pairwise dot-plots between all the sequences in the set. The dot-plots are first simplified by considering only the projections of the diagonal segments of similarity onto the axes. From these projections, a data matrix is built, in which each column is representative of the comparisons of one given sequence with all the other ones. This data matrix is then transformed into a distance matrix by a chi-squared anal., from which the coordinates of the sequences in an orthonormal Euclidean space are obtained. The sequences are finally classified by a dynamic clustering procedure followed by a search for strong clusters . Application of this method to protein families such as the globins, the cytochromes c and the aminoacyl-tRNA synthetases shows that it is quite effective in delineating subgroups that contain even distantly related sequences.

L9 ANSWER 113 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1993:442281 CAPLUS
DN 119:42281
TI The .beta. globin gene cluster of the prosimian primate Galago crassicaudatus: Nucleotide sequence determination of the 41-kb cluster and comparative sequence analyses
AU Tagle, Danilo A.; Stanhope, Michael J.; Siemieniak, David R.; Benson, Philip; Goodman, Morris; Slightom, Jerry L.
CS Sch. Med., Wayne State Univ., Detroit, MI, 48201, USA
SO Genomics (1992), 13(3), 741-60 CODEN: GNMCEP; ISSN: 0888-7543
DT Journal
LA English
AB The nucleotide sequence of the .beta. globin gene cluster of the prosimian Galago crassicaudatus has been detd. A total sequence spanning 41,101 bp contains and links together previously published sequences of the five galago .beta.-like globin genes (5'-.epsilon.-.gamma.-.psi..eta.- .delta.-.beta.-3'). A computer-aided search for middle interspersed repetitive sequences identified 10 LINE (L1) elements, including a 5' truncated repeat that is orthologous to the full-length L1 element found in the human .epsilon.-.gamma. intergenic region. SINE elements that were identified included one Alu type I repeat, four Alu type II repeats, and two methionine tRNA-derived Monomer (type III) elements. Alu type II and Monomer sequences are unique to the galago genome. Structural analyses of the cluster sequence reveals that it is relatively A + T rich (about 62%) and regions with high G + C content are assocd. primarily with globin coding regions. Comparative analyses with the .beta. globin cluster sequences of human, rabbit, and mouse reveal extensive sequence homologies in their genic regions, but only human, galago, and rabbit sequences share extensive intergenic sequence homologies. Divergence analyses of aligned intergenic and flanking sequences from orthologous human, galago, and rabbit sequences show a gradation in the rate of nucleotide sequence evolution along the cluster where sequences 5' of the .epsilon. globin gene region show the least sequence divergence and sequences just 5' of the .beta. globin gene region show the greatest sequence divergence.

L9 ANSWER 114 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1993:208913 CAPLUS
DN 118:208913
TI A novel method of protein sequence classification based on oligopeptide frequency analysis and its application to search for functional sites and to domain localization
AU Solovyev, V. V.; Makarova, K. S.
CS Inst. Cystol. Genet., Novosibirsk, 63090, Russia
SO CABIOS, Computer Applications in the Biosciences (1993), 9(1), 17-24 CODEN: COABER; ISSN: 0266-7061
DT Journal
LA English
AB A new method for distinguishing among protein families based on the anal. of oligopeptide compn. of amino acid sequences is presented. It is assumed that any protein family can be characterized by a set of essential oligopeptides (oligopeptide vocabulary). A simple approach to find such a vocabulary is suggested. It is shown that comparison of the vocabularies can distinguish among different families and the latter from random sequences. This comparison can be successfully made with a small set of frequencies of 25 dipeptides (or tripeptides). No preliminary alignment is necessary. It is established that characteristic peptides are located in the regions of functional value, as shown for GTP-binding domains of the translation elongation factors. It is demonstrated that this method is reasonably efficient for localizing functional domains in the amino acid sequences. The av. error of prediction does not exceed three or four amino acid residues as shown for several functional domains.

L9 ANSWER 115 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1993:162282 CAPLUS
DN 118:162282
TI Phylogenetic and structural analyses of MMTV LTR ORF sequences of exogenous and endogenous origins
AU Brandt-Carlson, Carolyn; Butel, Janet S.; Wheeler, David
CS Div. Mol. Virol., Baylor Coll. Med., Houston, TX, 77030, USA

SO Virology (1993), 193(1), 171-85 CODEN: VIRLAX; ISSN: 0042-6822
DT Journal
LA English
AB The long terminal repeat (LTR) of mouse mammary tumor virus (MMTV) harbors an open reading frame (ORF) that encodes a glycoprotein and is present in all exogenous and endogenous MMTV proviruses. The ORF protein has been reported to interact with the immune system of mice to cause deletion of specific V.beta.-bearing subsets of T cells. Twenty-two MMTV LTR ORF sequences were analyzed . Although highly conserved, the MMTV ORF sequences are not identical, with .apprxeq.35% of the total variation clustered at the carboxy terminus. Statistical anal. revealed the presence of 2 conserved regions in the protein , one of which contained a transmembrane-like domain (residues 45-63). Two potential nuclear localization signals were recognized. Many ORF sequences shared polymorphisms. To analyze relationships, phylogenetic trees were constructed on the basis of alignments of LTR ORF sequences. A tree generated from the carboxy-terminal 35 residues clustered the sequences into three divergent families. The topol. of the tree based on the N-terminal 288 residues differed significantly, with some MMTV sequences rearranged relative to their C-terminal families. A continuum of exogenous-like to endogenous-like character was suggested by the N-terminal tree. The discordance between the topologies of the 2 trees suggests that some type of genetic exchange has occurred in the MMTV LTR gene. Mechanisms and implications of such genetic exchange are discussed.

L9 ANSWER 116 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1993:35417 CAPLUS
DN 118:35417
TI Determination of the base recognition positions of zinc fingers from sequence analysis
AU Jacobs, Grant H.
CS Struct. Stud. Div., MRC Lab. Mol. Biol., Cambridge, CB2 2QH, UK
SO EMBO Journal (1992), 11(12), 4507-17 CODEN: EMJODG; ISSN: 0261-4189
DT Journal
LA English
AB The CC/HH zinc finger is a small independently folded DNA recognition motif found in many eukaryotic proteins , which ligates zinc through two cysteine and two histidine ligands. A database of 1340 zinc fingers from 221 proteins has been constructed and a program for anal. of aligned sequences written. This paper describes sequence anal. aimed at detg. the amino acid positions that recognize the DNA bases, by comparing two types of sequence variation. Using the idea that long runs of adjacent zinc fingers have arisen from internal gene duplication, the conservation of each position of the finger within the runs was calcd. The conservation of each position of the finger between homologous proteins from different species was also noted. A correlation of the two types of conservation showed clusters of related amino acids. One cluster of three positions was found to be esp. variable within long runs, but highly conserved between corresponding fingers of homologous proteins ; these positions are predicted to be the base contact positions. They match the amino acid positions that contact the bases in the cocrystal structure detd. by Pavletich and Pabo [Science, 240, 809-817 (1991)]. An adjacent cluster of four positions on the plot may also be assocd. with DNA binding. This anal. shows that the base recognition positions can be identified even in the absence of a known structure for a zinc finger. These results are applicable to zinc fingers where the structure of the complex is unknown, in particular suggesting that the individual finger-DNA interaction seen in the Zif268-DNA structure has been conserved in many zinc finger-DNA interactions.

L9 ANSWER 117 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1992:567126 CAPLUS
DN 117:167126
TI A comparison of several similarity indexes used in the classification of protein sequences : a multivariate analysis
AU Landes, Claudine; Henaut, Alain; Risler, Jean Loup
CS Cent. Genet. Mol., CNRS, Gif-sur-Yvette, 91198, Fr.
SO Nucleic Acids Research (1992), 20(14), 3631-7 CODEN: NARHAD; ISSN: 0305-1048
DT Journal
LA English
AB The present work describes an attempt to identify reliable criteria which could be used as distance indexes between protein sequences. Seven different criteria have been tested: i and ii) the scores of the alignments as given by the BESTFIT and the FASTA programs; iii) the ratio parameter, i.e. the BESTFIT score divided by the length of the aligned peptides ; iv and v) the statistical significance (Z-scores) of the scores provided by the program RELATE which performs a segment-by-segment comparison of 2 sequences, and vii) an original distance index calcd. by the program DOCMA from all the pairwise dotplots between the sequences. These 7 criteria have been tested against the amino acid sequences of 39 globins and those of the 20 aminoacyl-tRNA synthetases from E. coli. The distances between the sequences were analyzed by the multivariate anal. techniques. The results show that the distances calcd. from the scores of the pairwise alignments are not adequately sensitive. The Z-score from relate is not selective enough and too demanding in computer time. Three criteria gave a classification consistent with the known similarities between the sequences in the sets, namely the Z-scores from BESTFIT and FASTA and the multiple dotplot comparison distance index from DOCMA.

L9 ANSWER 118 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1992:546115 CAPLUS
DN 117:146115
TI Relationships, derived from optimum alignments , among amino acid sequences of plant peroxidases
AU Tyson, Hugh
CS Biol. Dep., McGill Univ., Montreal, QC, H3A 1B1, Can.
SO Canadian Journal of Botany (1992), 70(3), 543-56 CODEN: CJBOAW; ISSN: 0008-4026
DT Journal
LA English
AB The amino acid and (or) DNA sequences of 13 plant peroxidases (EC 1.11.1.7), which include isoenzymes within species, are currently available in data bases; all have similar lengths of approx. 300 amino acids. Sequence relationships among these 13, plus 2 microbial peroxidases of similar length, were examd. The 15 sequences were compared in all 105 pairwise combinations using optimum alignment procedures. Gap penalties were detd. from anal. of penalty change effects. Distances between sequences generated by optimum alignments were analyzed by clustering techniques to generate effects. Distances between sequences, which provided pairwise distance measurements independent of the av. distance for a sequence, were used to evaluate sequence similarities; closely related sequences produce closely correlated specific distances. Among the seven plant species, five subgroups were established: (1) horseradish isoperoxidases, (2) turnip and wheat, (3) cucumber and

tobacco, (4) potato and tomato, and (5) in which cytochrome c peroxidase showed some similarity to ligninase, but both were only distantly related to plant peroxidase. Horseradish isoperoxidases were related to sequences in subgroups 2, 3, and 4 but resembled subgroups 2 and 3 more closely than 4. Subgroup 2 was more related to 3 than any other.

L9 ANSWER 119 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1992:77590 CAPLUS
DN 116:77590
TI Sequence and comparative analysis of the rabbit .alpha.-like globin gene cluster reveals a rapid mode of evolution in a G + C-rich region of mammalian genomes
AU Hardison, Ross; Krane, Dan; Vandenbergh, David; Cheng, Jan Fang; Mansberger, James; Taddie, John; Schwartz, Scott; Huang, Xiaoqiu; Miller, Webb
CS Inst. Mol. Evol. Genet., Pennsylvania State Univ., University Park, PA, 16802, USA
SO Journal of Molecular Biology (1991), 222(2), 233-49
CODEN: JMOBAK; ISSN: 0022-2836
DT Journal
LA English
AB A sequence of 10,621 base-pairs from the .alpha.-like globin gene cluster of rabbit was detd. It includes the sequence of gene .zeta.1 (a pseudogene for the rabbit embryonic .zeta.-globin), the functional rabbit .alpha.-globin gene, and the .theta.1 pseudogene, along with the sequences of eight C repeats (short interspersed repeats in rabbit) and a J sequence implicated in recombination. The region is quite G + C-rich (62%) and contains two CpG islands. As expected for a very G + C-rich region, it has an abundance of open reading frames, but few of the long open reading frames are assocd. with the coding regions of genes. Alignments between the sequences of the rabbit and human .alpha.-like globin gene clusters reveal matches primarily in the immediate vicinity of genes and CpG islands, while the intergenic regions of these gene clusters have many fewer matches than are seen between the .beta.-like globin gene clusters of these two species. Furthermore, the non-coding sequences in this portion of the rabbit .alpha.-like globin gene cluster are shorter than in human, indicating a strong tendency either for sequence contraction in the rabbit gene cluster or for expansion in the human gene cluster . Thus, the intergenic regions of the .alpha.-like globin gene clusters have evolved in a relatively fast mode since the mammalian radiation, but not exclusively by nucleotide substitution. Despite this rapid mode of evolution, some strong matches are found 5' to the start sites of the human and rabbit .alpha. genes, perhaps indicating conservation of a regulatory element. The rabbit J sequence is over 1000 base-pairs long; it contains a C repeat at its 5' end and an internal region of homol. to the 3'-untranslated region of the .alpha.-globin gene. Part of the rabbit J sequence matches with sequences within the X homol. block in human. Both of these regions have been implicated as hot-spots for recombination, hence the matching sequences are good candidates for such a function. All the interspersed repeats within both gene clusters are retroposon SINEs that appear to have inserted independently in the rabbit and human lineages.

L9 ANSWER 120 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1991:627036 CAPLUS
DN 115:227036
TI Evolution and relatedness in two aminoacyl-tRNA synthetase families
AU Nagel, Glenn M.; Doolittle, Russell F.
CS Cent. Mol. Genet., Univ. California, San Diego, La Jolla, CA, 92003, USA
SO Proceedings of the National Academy of Sciences of the United States of America (1991), 88(18), 8121-5 CODEN: PNASA6; ISSN: 0027-8424
DT Journal
LA English
AB Sequence segments of about 140 amino acids in length, each contg. a selected consensus region, were used in alignments of the aminoacyl-tRNA synthetases with the aim of discerning their evolutionary relationships. In all cases tested, enzymes specific for the same amino acid from a variety of organisms grouped together, reinforcing the supposition that the aminoacyl-tRNA synthetases are very ancient enzymes that evolved to include the full complement of 20 amino acids long before the divergence leading to prokaryotes and eukaryotes. The enzymes are divided into two mutually exclusive groups that appear to have evolved from independent roots. Group I, for which two sequence segments were analyzed , contains the enzymes specific for glutamic acid, glutamine, tryptophan, tyrosine, valine, leucine, isoleucine, methionine, and arginine. Group II enzymes include those activating threonine, proline, serine, lysine, aspartic acid, asparagine, histidine, alanine, glycine, and phenylalanine. Both groups contain a spectrum of amino acid types, suggesting the possibility that each could have once supported an independent system for protein synthesis. Within each group, enzymes specific for chem. similar amino acids tend to cluster together, indicating that a major theme of synthetase evolution involved the adaptation of binding sites to accommodate related amino acids with subsequent specialization to a single amino acid. In a few cases, however, synthetases activating dissimilar amino acids are grouped together.

L9 ANSWER 121 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1991:444500 CAPLUS
DN 115:44500
TI Amino acid sequence analysis of the annexin super-gene family of proteins
AU Barton, Geoffrey J.; Newman, Richard H.; Freemont, Paul S.; Crumpton, Michael J.
CS Lab. Mol. Biophys., Univ. Oxford, Oxford, OX1 3QU, UK
SO European Journal of Biochemistry (1991), 198(3), 749-60
CODEN: EJBCAI; ISSN: 0014-2956
DT Journal
LA English
AB Twenty-two available annexin sequences consisting of 88 similar repeat units were drawn together. Multiple sequence alignment , pattern matching, secondary structure prediction, and conservation anal. were used to characterize the mols. The anal. clearly shows that the repeats cluster into 4 distinct families and that greatest variation occurs within the repeat 3 units. Multiple alignment of the 88 repeats shows amino acids with conserved physicochem. properties at 22 positions, with only Gly-23 being absolutely conserved in all repeats. Secondary structure prediction techniques identify 5 conserved helixes in each repeat unit and patterns of conserved hydrophobic amino acids are consistent with 1 face of a helix packing against the protein core in predicted helixes a, c, d, e. Helix b is generally hydrophobic in all repeats, but contains a striking pattern of repeat-specific residue conservation at position 31, with arginine in repeats 4 and glutamate in repeats 2, but unconserved amino acids in repeats 1 and 3. This suggests repeats 2 and 4 may interact via a buried salt-bridge. The loop between predicted helixes a and b of repeat 3 shows features distinct from the equiv. loop in repeats 1, 2

and 4, suggesting an important structural and/or functional role for this region. No compelling evidence emerges from this study of uteroglobin and the annexins sharing similar tertiary structures, or for uteroglobin representing a deriv. of a primordial 1-repeat structure that underwent duplication to give the present day annexins. The analyses performed in this paper are re-evaluated in the Appendix, in the light of the recently published x-ray structure for human annexin V. The structure confirms most of the predictions and shows the power of techniques for the detn. of tertiary structural information from the amino acid sequences of an aligned protein family.

L9 ANSWER 122 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1991:58455 CAPLUS
DN 114:58455
TI Hydrophobic cluster analysis: procedures to derive structural and functional information from 2-D-representation of protein sequences
AU Lemesle-Varloot, L.; Henrissat, B.; Gaboriaud, C.; Bissery, V.; Morgat, A.; Mornon, J. P.
CS Lab. Mineral.-Cristallogr., Univ. Paris, Paris, 75252, Fr.
SO Biochimie (1990), 72(8), 555-74 CODEN: BICMBE; ISSN: 0300-9084
DT Journal; General Review
LA English
AB Hydrophobic cluster anal. (HCA) (Gaboriaud, C. et al., 1987) is a very efficient method to analyze and compare protein sequences. Despite its effectiveness, this method is not widely used because it relies in part on the experience and training of the user. Detailed guidelines as to the use of HCA are presented and include discussions on: the definition of the hydrophobic clusters and their relationships with secondary and tertiary structures; the length of the clusters ; the amino acid classification used for HCA; the HCA plot programs; and the working strategies. Various procedures for the anal. of a single sequence are presented: structural segmentation, structural domains and secondary structure evaluation. Like most sequence anal. methods, HCA is more efficient when several homologous sequences are compared. Procedures for the detection and alignment of distantly related proteins by HCA are described through several published examples along with 2 previously unreported cases: the .beta.-glucosidase from Ruminococcus albus is clearly related to the .beta.-glucosidases from Clostridium thermocellum and Hansenula anomala although they display a reverse organization of their constitutive domains; the alignment of the sequence of human GTPase activating protein with that of the Crk oncogene is presented. Finally, the pertinence of HCA in the identification of important residues for structure/function as well as in the prepn. of homol. modeling is discussed.

L9 ANSWER 123 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1990:625287 CAPLUS
DN 113:225287
TI Sequence analysis of the phosphoprotein (P) genes of human parainfluenza type 4A and 4B viruses and RNA editing at transcript of the P genes: the number of G residues added is imprecise
AU Kondo, Kunio; Bando, Hisanori; Tsurudome, Masato; Kawano, Mitsuo; Nishio, Machiko; Ito, Yasuhiko
CS Sch. Med., Mie Univ., Tsu, 514, Japan
SO Virology (1990), 178(1), 321-6 CODEN: VIRLAX; ISSN: 0042-6822
DT Journal
LA English
AB The authors cloned and sequenced the cDNAs against genomic RNAs and mRNAs for phosphoproteins (Ps) of human parainfluenza virus types 4A (PIV-4A) and 4B (PIV-4B). The PIV-4A and -4B P genes were 1535 nucleotides including poly(A) tract and were found to have 2 small open reading frames, neither of which was apparently large enough to encode the P protein . A cluster of G residues was found in genomic RNA, and the no. of G residues was 6 in both PIV-4A and -4B. However, the no. of G residues at the corresponding site in the mRNAs to the genomic RNA was not const. Three different mRNA cDNA clones were obtained; the first type of mRNA encodes a larger (P) protein of 399 amino acids, the second type encodes V protein of 229 or 230 amino acids, and the third type encodes the smallest protein (156 amino acids). Comparisons on the nucleotide and the amino acid sequences of P and V proteins between these 2 subtypes revealed extensive homologies. However, these homol. degrees are lower than that of NP protein . The C-terminal regions of the P and V proteins of PIV-4s could be aligned with all other paramyxoviruses, PIV-2, mumps virus (MuV), simian virus 5 (SV 5), Newcastle disease virus (NDV), measles virus (MV), canine distemper virus (CDV), Sendai virus (SV), and PIV-3. On the other hand, the P-V common (N-terminal) regions showed no homol. with MV, CDV, SV, and PIV-3. Seven phylogenetic trees of Paramyxoviruses were constructed from the entire and partial regions of P and V proteins .

L9 ANSWER 124 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1990:568012 CAPLUS
DN 113:168012
TI Amino acid compositions and partial sequences of two types of alkaline serine proteases from Nocardiopsis dassonvillei subsp. prasina OPC-210
AU Tsujiba, Hiroshi; Miyamoto, Katsushiro; Hasegawa, Toru; Inamori, Yoshihiko
CS Osaka Univ. Pharm. Sci., Matsubara, 580, Japan
SO Agricultural and Biological Chemistry (1990), 54(8), 2177-9 CODEN: ABCHA6; ISSN: 0002-1369
DT Journal
LA English
AB The alkalophilic actinomycete, N. dassonvillei prasina OPC-210, produces 2 types of alk. serine proteases (NDP-I and NDP-II). The purifn. and properties of these proteases, as well as the taxonomy of the alkalophilic actinomycete were previously reported. Here, the amino acid compns. and partial amino acid sequences of NDP-I and NDP-II isolated from the culture filtrate of N. dassonvillei prasina OPc-210 are reported. The amino acid compns. of NDP-I and NDP-II were detd. and compared with those of other microbial proteases. NDP-I contained 6 cysteine residues, which probably formed 3 disulfide bonds. The amino acid compn. of NDP-I was similar to that of Streptomyces griseus proteases A and B and .alpha.-lytic protease. On the other hand, NDP-II did not contain cysteine residues like subtilisins. The N-terminal 41 amino acid residues of NDP-I were sequenced and compared with those of other microbial serine proteases. From the alignment of these sequences, the partial N-terminal sequence of NDP-I showed a striking similarity to those of chymotrypsin-like proteases, i.e., S. griseus proteases A and B, S. griseus alk. protease, and .alpha.-lytic protease. From these results, NDP-I was classified as a chymotrypsin-type serine protease. The N-terminal amino acid sequence of NDP-II was analyzed and compared with those of subtilisin BPN, elastase YaB, thermitase, proteinase K, and aqualysin 1. Surprisingly, the partial amino acid sequence of NDP-II showed striking homol. with that of aqualysin I (65% homol.). This is the 1st reported

example of an aqualysin I-like alk. serine protease produced by an alkalophilic actinomycete.

L9 ANSWER 125 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1990:472168 CAPLUS
DN 113:72168
TI Cloning and structural characterization of porcine heart aconitase
AU Zheng, Limin; Andrews, P. C.; Hermodson, Mark A.; Dixon, Jack E.; Zalkin, Howard
CS Dep. Biochem., Purdue Univ., West Lafayette, IN, 47907, USA
SO Journal of Biological Chemistry (1990), 265(5), 2814-21
CODEN: JBCHA3; ISSN: 0021-9258
DT Journal
LA English
AB A full-length cDNA encoding porcine heart aconitase was derived from .lambda.gt10 recombinant clones and by amplification of the 5' end of the mRNA. The 2700 bp cDNA contains a 29-bp 5' untranslated region, a 2343-bp coding segment, and a 327-bp 3' untranslated region. The porcine heart enzyme is synthesized as a precursor contg. a mitochondrial targeting sequence of 27 amino acid residues which is cleaved to yield a mature enzyme of 754 amino acids, Mr = 82,754, having a blocked amino terminus. The NH2-terminal pyroglutamyl residue of the mature enzyme was identified by fast atom bombardment mass spectrometry and sequence analyses of an NH2-terminal peptide . Mature porcine heart aconitase contains 12 cysteine residues. An alignment of the derived porcine heart sequence with 8 cysteine-contg. tryptic peptides from bovine heart aconitase shows that 198 of 202 amino acids are conserved and suggests that the 2 enzymes are virtually identical. Cysteines 358, 421, and 424 are ligands to the Fe-S cluster in the inactive [3Fe-4S] and active [4Fe-4S] forms. An alignment of the derived porcine heart sequence with 8 cysteine-contg. tryptic peptides from bovine heart aconitase shows that 198 of 202 amino acids are conserved and suggests that the 2 enzymes are virtually identical.

L9 ANSWER 126 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1989:527821 CAPLUS
DN 111:127821
TI Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles Halobacterium halobium and Halococcus morrhuae
AU Leffers, Henrik; Gropp, Felix; Lottspeich, Friedrich; Zillig, Wolfram; Garrett, Roger A.
CS Kem. Inst., Aarhus Univ., Aarhus, DK-8000, Den.
SO Journal of Molecular Biology (1989), 206(1), 1-17 CODEN: JMOBAK; ISSN: 0022-2836
DT Journal
LA English
AB The genes for the 4 largest subunits, A, B', B" and C, of the DNA-dependent RNA polymerase were cloned from the extreme halophile H. halobium and sequenced , and their transcription was analyzed . The downstream half of this gene cluster from another extreme halophile, H. morrhuae, was also cloned and sequenced and its transcription products were characterized. The H. halobium genes were transcribed into a common transcript from an upstream promoter in the order B", B', A and C. They are flanked by, and co-transcribed with, two smaller genes coding for 75 and 139 amino acid residues, resp. Immediately downstream from these genes were two open reading frames that are homologous to ribosomal proteins S12 and S7 from Escherichia coli. In both extreme halophiles, these genes were transcribed from their own promoter, but in H. morrhuae there was also considerable read-through from the RNA polymerase genes. Sequence alignment studies showed that the combined B" + B' subunits are equiv. to the B subunits of the eukaryotic polymerases I and II and to the eubacterial .beta. subunit, while the combined A+C subunits correspond to the A subunits of eukaryotic RNA polymerases I, II, and III and to the eubacterial .beta.' subunit. The sequence similarity to the eukaryotic subunits was always much higher than to the eubacterial subunits. Conserved sequence regions within the individual subunits were located which are likely to constitute functionally important domains; they include sites assocd. with rifampicin and .alpha.-amanitin binding and two possible zinc binding fingers. Phylogenetic analyses based on sequence alignments confirmed that the extreme halophiles belong to the archaebacterial kingdom.

L9 ANSWER 127 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1989:473604 CAPLUS
DN 111:73604
TI Primary structure of a member of the serpin superfamily of proteinase inhibitors from an insect, Manduca sexta
AU Kanost, Michael R.; Prasad, Sarvamangala V.; Wells, Michael A.
CS Dep. Biochem., Univ. Arizona, Tucson, AZ, 85721, USA
SO Journal of Biological Chemistry (1989), 264(2), 965-72
CODEN: JBCHA3; ISSN: 0021-9258
DT Journal
LA English
AB A cDNA clone isolated from a fat body cDNA library from M. sexta was sequenced and shown to code for a member of the serpin family of proteinase inhibitors. The cDNA had an open reading frame which coded for a 392-residue polypeptide of mol. wt. 43,500 with a hydrophobic N-terminal sequence which appeared to be a signal peptide . An alignment of this amino acid sequence with 11 members of the serpin superfamily revealed that the insect protein was 25-30% identical with most members of the superfamily. The alignment was used to construct an evolutionary tree of the serpin sequences analyzed , which indicated that the progenitor of the M. sexta serpin and the human serpins most closely related to it diverged from other serpin genes prior to the divergence of the vertebrates and invertebrates. The M. sexta serpin was predicted to inhibit elastase due to the presence of alanine at the P1 position of its reactive center and was classified as an alaserpin. A glycoprotein of mol. wt. 47,000 isolated from hemolymph of M. sexta larvae had an N-terminal sequence identical to that deduced from the alaserpin cDNA clone and inhibited porcine pancreatic elastase and bovine chymotrypsin.

L9 ANSWER 128 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1988:217949 CAPLUS
DN 108:217949
TI Primary- and secondary-structural analysis of a unique prokaryotic metallothionein from a Synechococcus sp. cyanobacterium
AU Olafson, Robert W.; McCubbin, William D.; Kay, Cyril M.
CS Dep. Biochem. Microbiol., Univ. Victoria, Victoria, BC, V8W 2Y2, Can.
SO Biochemical Journal (1988), 251(3), 691-9 CODEN: BIJOAK; ISSN: 0306-3275
DT Journal
LA English
AB Biochem. and physiol. studies of Synechococcus cyanobacteria have indicated the presence of a low-mol.-wt. (Mr) heavy-metal-binding protein with marked similarity to

eukaryotic metallothioneins (MTs). The characterization of a Synechococcus prokaryotic MT isolated by gel-permeation and reverse-phase chromatog is reported . The large no. of variants of this mol. found during chromatog. sepn. could not be attributed to the presence of major isoproteins as assessed by amino acid anal. and amino acid sequencing of isoforms. Two of the latter had identical primary structures that differed substantially from the well-described eukaryotic MTs. In addn. to 6 long-chain aliph. residues, 2 arom. residues were found adjacent to one another near the center of the mol., making this the most hydrophobic MT to be described. Other unusual features included a pair of histidine residues located in repeating Gly-His-Thr-Gly sequences near the C-terminus and a complete lack of assocn. of hydroxylated residues with cysteine residues, as is commonly found in eukaryotes. Similarly, aside from a single lysine residue, no basic amino acid residues were found adjacent to cysteine residues in the sequence. Most importantly, sequence alignment analyses with mammalian, invertebrate and fungal MT sequences showed no statistically significant homol. aside from the presence of Cys-Xaa-Cys (Xaa = amino acid) structures common to all MTs. On the other hand, like other MTs, the prokaryotic mol. appears to be free of .alpha.-helical structure but has a considerable amt. of .beta.-structure, as predicted by both CD measurements and the Chou and Fasman empirical relations. Considered together, these data suggested that some similarity between the metal-thiolate clusters of the prokaryote and eukaryote MTs may exist.

L9 ANSWER 129 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1988:51977 CAPLUS
DN 108:51977
TI Primary structure homologies between two zinc metallopeptidases, the neutral endopeptidase 24.11 ("enkephalinase") and thermolysin, through clustering analysis
AU Benchetrit, T.; Bissery, V.; Mornon, J. P.; Devault, A.; Crine, P.; Roques, B. P.
CS Dep. Chim. Org., INSERM, Paris, 75006, Fr.
SO Biochemistry (1988), 27(2), 592-6 CODEN: BICHAW; ISSN: 0006-2960
DT Journal
LA English
AB Analogies in the sequences of 2 related Zn-contg. metalloproteinases, thermolysin (I) (316 amino acids) and the recently cloned membrane metalloendopeptidase (neutral endopeptidase 24.11, enkephalinase) (II) (749 amino acids) were demonstrated by use of a hydrophobic cluster anal. method derived from the theory of V. I. Lim (1974). Two sequence alignments were proposed for the entire primary structure of I and the C-terminal part of II. Except for an arginine residue, all of the amino acids involved in the active site of I were retrieved in both models of II within conserved clustered structures. The 1st model was characterized by a deletion of the Ca2+-binding coil present in I and the 2nd by replacement of this coil by 2 .alpha.-helixes. In both models an arginine residue could be located in the active site of II.

L9 ANSWER 130 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1987:401780 CAPLUS
DN 107:1780
TI Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine
AU Altenbach, Susan B.; Pearson, Karen W.; Leung, Filomena W.; Sun, Samuel S. M.
CS Plant Cell Res. Inst., ARCO, Dublin, CA, 94568, USA
SO Plant Molecular Biology (1987), 8(3), 239-50 CODEN: PMBIDB; ISSN: 0167-4412
DT Journal
LA English
AB The primary amino acid sequence of an abundant methionine-rich seed protein found in Brazil nut (Bertholletia excelsa H.B.K.) was elucidated by protein sequencing and from the nucleotide sequence of cDNA clones. The 9 kDa subunit of this protein contained 77 amino acids, of which 14 were methionine (18%) and 6 were cysteine (8%). Over half of the methionine residues in this subunit are clustered in two regions of the polypeptide , where they are interspersed with arginine residues. In one of these regions, methionine residues account for 5 out of 6 amino acids, and 4 of these methionine residues are contiguous. The sequence data verifies that the Brazil nut sulfur-rich protein is synthesized as a precursor polypeptide that is considerably larger than either of the 2 subunits of the mature protein . Three proteolytic processing steps by which the encoded polypeptide is sequentially trimmed to the 9 kDa and 3 kDa subunit polypeptides were correlated with the sequence information. The sulfur-rich protein from Brazil nut is homologous in its amino acid sequence to small water-sol. proteins found in 2 other oilseeds, castor bean (Ricinus communis) and rapeseed (Brassica napus). When the amino acid sequences of these 3 proteins are aligned to maximize homol., the arrangement of cysteine residues is conserved. However, the 2 subunits of the Brazil nut protein contain over 19% methionine, whereas the homologous proteins from castor bean and rapeseed contain only 2.1% and 2.6% methionine, resp.

L9 ANSWER 131 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1986:203439 CAPLUS
DN 104:203439
TI The classification of amino acid conservation
AU Taylor, William Ramsay
CS Dep. Crystallogr., Birkbeck Coll., London, WC1E 7HX, UK
SO Journal of Theoretical Biology (1986), 119(2), 205-18, 1 plate CODEN: JTBIAP; ISSN: 0022-5193
DT Journal
LA English
AB A classification of amino acid type is described which is based on a synthesis of physicochem. and mutation data. This is organized in the form of a Venn diagram from which subsets are derived that include groups of amino acids likely to be conserved for similar structural reasons. These sets are used to describe conservation in aligned sequences by allocating to each position the smallest set that contains all the residue types brought together by alignment . This minimal set assignment provides a simple way of reducing the information contained in a sequence alignment to a form which can be analyzed by computer yet remains readable.

L9 ANSWER 132 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1985:574603 CAPLUS
DN 103:174603
TI Replacement by site-directed mutagenesis indicates a role for histidine 170 in the glutamine amide transfer function of anthranilate synthase
AU Amuro, Naoki; Paluh, Janet L.; Zalkin, Howard
CS Dep. Biochem., Purdue Univ., West Lafayette, IN, 47907, USA
SO Journal of Biological Chemistry (1985), 260(27), 14844-9 CODEN: JBCHA3; ISSN: 0021-9258
DT Journal
LA English
AB Anthranilate synthase is a glutamine amidotransferase that catalyzes the 1st reaction in tryptophan biosynthesis. Conserved amino acid residues likely to be essential for

glutamine-dependent activity were identified by alignment of the glutamine amide transfer domains in 4 different enzymes: anthranilate synthase component II (AS II), p-aminobenzoate synthase component II, GMP synthetase, and carbamoyl phosphate synthetase. Conserved amino acids were mainly localized in 3 clusters . A single conserved histidine (His), AS II His-170, was replaced by tyrosine (Tyr) by using site-directed mutagenesis. Glutamine-dependent enzyme activity was undetectable in the Tyr-170 mutant, whereas the NH3-dependent activity was unchanged. Affinity labeling of AS II active site cysteine (Cys)-84 by 6-diazo-5-oxonorleucine was used to distinguish whether His-170 has a role in formation or in breakdown of the covalent glutaminyl-Cys-84 intermediate. His-170 appears to function as a general base to promote glutaminylation of Cys-84. Reversion anal. was consistent with a proposed role of His-170 in catalysis as opposed to a structural function. These expts. demonstrate the application of combining sequence analyses to identify conserved, possibly functional amino acids, site-directed mutagenesis to replace candidate amino acids, and protein chem. for anal. of mutationally altered proteins , a regimen that can provide new insights into enzyme function.

L9 ANSWER 133 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1985:519344 CAPLUS
DN 103:119344
TI Extraction of symbolic patterns common to a family of biosequences
AU Saurin, William; Marliere, Philippe
CS Unite Programmation Mol. Toxicol. Genet., Inst. Pasteur, Paris, 75015, Fr.
SO Biochimie (1985), 67(5), 517-21 CODEN: BICMBE; ISSN: 0300-9084
DT Journal
LA French
AB A method and programs were developed for extg. symbolic patterns for elucidating the sequence of biol. macromols. such as proteins and nucleic acids. A set of sequences can be defined by their common subsequences, and the length of these is a measure of the overall resemblance of the set. Each subsequence corresponds to a succession of symbols embedded in every sequence, following the same order but not necessarily contiguous. Detg. the longest common subsequence (LCS) requires the exhaustive testing of all possible common subsequences, which sum up to about 2L, if L is the length of the shortest sequence. A polynomial algorithm (O(n.L4) is presented where n is the no. of sequences) for generating strings related to the LCS and constructed with the sequence alphabet and an indetn. symbol. Such strings are iteratively improved by deleting indetn. symbols and concomitantly introducing the greatest no. of alphabet symbols. Processed accordingly, nucleic acid and protein sequences lead to keywords encompassing the salient positions of homologous chains, which can be used for aligning or classifying them, as well as for finding related sequences in data banks. Examples are given of the application of the method to extg. anchorage points of Escherichia coli tRNA sequences and for the recognition of indistinct determinants in translation initiation sites of E. coli genes.

L9 ANSWER 134 OF 134 CAPLUS COPYRIGHT 2003 ACS
AN 1982:157667 CAPLUS
DN 96:157667
TI Primary structure of the polypeptide chain elongation factor Tu from E. coli. I. Amino acid sequence of fragment B
AU Nakamura, Shun; Nakayama, Naoki; Takahashi, Kenji; Kaziro, Yoshito
CS Inst. Med. Sci., Univ. Tokyo, Tokyo, 108, Japan
SO Journal of Biochemistry (Tokyo, Japan) (1982), 91(3), 1047-63 CODEN: JOBIAO; ISSN: 0021-924X
DT Journal
LA English
AB The complete amino acid sequence of fragment B obtained by the limited tryptic digestion of Escherichia coli polypeptide chain elongation factor Tu (EF-Tu) was detd. Seven peptides formed from fragment B by cleavage with CNBr (designated as CB1 to CB7 according to their order of alignment from N- to C-termini of fragment B) were purified, and 6 of them were completely sequenced by the manual method of sequential Edman degrdn. with direct identification of the phenylthiohydantoin amino acids. The remaining CNBr peptide (CB6), contg. 109 amino acid residues, was further digested with trypsin. Twelve tryptic peptides (designated as T1 to T12 according to their order of alignment from N- to C-termini of CB6) were isolated, and their amino acid sequences were analyzed . The alignment of CB peptides was based on the results of the automated sequence anal. of fragment B from its N-terminus and the sequence anal. of the overlapping peptides contg. SH groups obtained by the complete tryptic digestion of fragment B. The alignment of peptides T1 to T12 on CB6 was based on the result of the automated sequence anal. of CB6 and the sequence of the overlapping peptide obtained by the chem. cleavage of CB6 at the tryptophan residue using CNBr in heptafluorobutyric acid. The nucleotide sequence of the tufA gene was also utilized for the alignment of these peptides . Fragment B comprises amino acid residues 59-263 of E. coli EF-Tu, which consists of 393 amino acids. It contains the 2 functional and 1 nonfunctional SH groups of EF-Tu. All of the 5 histidine residues in fragment B were distributed within the 1st N-terminal quarter, and 3 of them were clustered around 1 of the functional SH groups. Although E. coli EF-Tu consists of 2 gene products (tufA and tufB), there was no microheterogeneity in the amino acid sequence of fragment B.

=> log y
COST IN U.S. DOLLARS SINCE FILE TOTAL ENTRY SESSION
FULL ESTIMATED COST 356.69 356.90

DISCOUNT AMOUNTS (FOR QUALIFYING ACCOUNTS) SINCE FILE TOTAL ENTRY SESSION
CA SUBSCRIBER PRICE -87.23 -87.23

STN INTERNATIONAL LOGOFF AT 12:39:15 ON 02 MAY 2003